DOCUMENT PUBLIC

Prise en compte de l'incertitude dans l'évaluation du risque d'exposition aux polluants du sol

Programme GESSOL

Rapport final

Juin 2002
BRGM/RP-51683-FR

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE DE L'ÉCOLOGIE ET DU DÉVELOPPEMENT DURABLE

DOCUMENT PUBLIC

Prise en compte de l'incertitude dans l'évaluation du risque d'exposition aux polluants du sol

Programme GESSOL

Rapport final

D. Guyonnet, B. Bourgine, J.-P. Chilès

Avec la collaboration de :
B. Côme, D. Dubois, H. Fargier

Juin 2002
BRGM/RP-51683-FR

BRGM

Mots clés : Incertitude, Evaluation des risques, Plantes, Métaux.

En bibliographie, ce rapport sera cité de la façon suivante :

Guyonnet, D., Bourgine, B., Chilès, J.-P., Dubois, D., Fargier, H., Côme, B. (2002) – Prise en compte de l'incertitude dans l'évaluation du risque d'exposition aux polluants du sol. Programme GESSOL. Rapport final. Rapport BRGM RP-51683-FR

© BRGM, 2002, ce document ne peut être reproduit en totalité ou en partie sans l'autorisation expresse du BRGM.

Synthèse

Dans le cadre du Programme GESSOL, piloté par l'INRA pour le Ministère chargé de l'Environnement, le BRGM a réalisé un projet relatif à la prise en compte de l'incertitude dans l'évaluation du risque d'exposition de l'homme aux polluants du sol. Le principal objectif de ce projet est de développer une méthodologie permettant de prendre en compte l'incertitude liée aux valeurs des paramètres qui influencent le risque d'exposition, d'une manière qui soit cohérente par rapport à l'information dont on dispose dans la pratique.

Comme on s'intéresse ici à l'exposition de l'homme aux polluants du sol par le biais de la consommation de légumes, la première partie du rapport présente une synthèse sur l'assimilation des métaux par les plantes. Cette synthèse souligne la difficulté de prévoir le transfert sol - plante, en raison de la multiplicité des facteurs d'influence (spéciation du métal, type de sol, type de plante, etc.). Il est donc préférable, dans la pratique, de disposer de mesures directes de l'assimilation de métaux par des plantes cultivées sur site.

Deux principaux modes de représentation de l'incertitude liée aux valeurs des paramètres influençant le risque d'exposition sont présentés : les modes « probabiliste » qui utilise des fonctions de distribution de probabilité, et « possibiliste » qui fait appel à des nombres dits flous. L'approche possibiliste est particulièrement adaptée lorsque l'information dont on dispose est partielle ou vague. Or dans la pratique, il est fréquent que l'on ait affaire à la fois aux deux types d'information. Une méthode, dite « hybride », est proposée pour combiner ces deux modes de représentation de l'incertitude dans une même évaluation du risque d'exposition. Cette méthode, qui associe la technique dite Monte Carlo au calcul en nombres flous, permet d'éviter d'introduire artificiellement de la connaissance en supposant a priori des distributions de probabilité pour des paramètres pour lesquels on ne dispose que de peu d'information.

La méthode hybride est appliquée à un cas réel de site industriel dont les sols superficiels présentent des teneurs élevées en métaux lourds. L'application illustre le potentiel de cette méthode, qui permet d'exploiter l'information « riche » fournie par les distributions de probabilité, tout en conservant le caractère sécuritaire du calcul en nombres flous. Elle présente donc un intérêt en terme d'une application « raisonnable » du principe de précaution.

L'application concrète est suivie par un traitement du problème dit « inverse » : pour une possibilité de dépassement de la dose jugée tolérable, et compte tenu de l'incertitude liée aux paramètres du modèle, quelle peut être la concentration résiduelle maximale dans le sol (après un éventuel traitement) ?

Sommaire

1. Introduction **5**

2. Synthèse sur les modèles d'assimilation des métaux par les plantes **6**

2.1. Introduction **6**

2.2. Les facteurs qui peuvent influencer le transfert sol-plante **9**

2.3. Les modèles de transfert sol-plante **10**

2.3.1. Introduction 10

2.3.2. Modèle A : HESP 11

2.3.3. Modèle B 13

2.3.4. Comparaison avec des données mesurées 15

2.3.5. Conclusions sur les modèles d'assimilation des métaux par les plantes 18

3. Prise en compte de l'incertitude dans l'évaluation du risque d'exposition **19**

3.1. Introduction **19**

3.2. Représentations probabiliste et possibiliste de l'incertitude **19**

3.3. Calcul Monte Carlo et calcul en nombres flous **22**

3.4. Comparaisons avec des critères de tolérance **25**

3.5. Développement d'une approche « hybride » probabiliste-possibiliste **28**

4. Application au cas d'un sol pollué par du cadmium **33**

4.1. Données mesurées **33**

4.2. Analyse géostatistique **35**

4.3. Analyse corrélatoire floue **38**

4.4. Calcul du risque d'exposition et représentation graphique **41**

4.5. Conséquences de l'hypothèse de distributions de probabilité a priori **43**

5. Traitement du problème « inverse » **47**

5.1. Expression du problème inverse **47**

5.2. Méthodologie générale **47**

5.3. Méthodologie spécifique **48**

5.4. Discussion **49**

6. Conclusions et perspectives **53**

Annexe 1 – Synthèse sur la combinaison des représentations probabiliste et possibiliste **59**

Annexe 2 – Article dans le *Journal of Environmental Engineering* (sous presse) **83**

1. Introduction

Le Programme GESSOL, piloté par l'INRA pour le Ministère chargé de l'Environnement, vise notamment à définir des critères de qualité des sols qui ne soient pas uniquement liés à des concepts de fertilité physico-chimique et biologique, mais qui intègrent également les aspects liés à la santé humaine. L'évaluation d'un effet sur la santé humaine est généralement effectuée en intégrant, d'une part, le risque d'exposition d'une cible humaine aux polluants du sol, et d'autre part, une connaissance des relations doses-effets. Cette évaluation fait intervenir de nombreux paramètres dont la plupart sont, dans la pratique, entachées d'une certaine incertitude. Cette incertitude peut avoir essentiellement deux origines. Elle peut soit être liée au fait que le paramètre présente un caractère aléatoire (en raison de la variabilité naturelle du phénomène en jeu), ou alors elle peut être due à la nature partielle et incomplète de l'information disponible.

Comme il sera montré dans ce rapport, différents *types* d'incertitude appellent différents *modes* de représentation de l'incertitude. Une question scientifique importante est donc : **comment peut-on combiner différents modes de représentation de l'incertitude dans une même évaluation du risque d'exposition de l'homme aux polluants du sol ?** Le principal objectif du projet de recherche proposé par le BRGM dans le cadre de GESSOL est de développer une méthodologie permettant d'effectuer cette combinaison. Une telle méthodologie permet dès lors de prendre en compte l'incertitude dans l'évaluation du risques d'exposition, d'une manière qui est **cohérente avec l'information dont on dispose dans la pratique**.

Ce rapport est divisé en 6 chapitres. Le scénario d'exposition qui nous intéresse dans le cadre de ce projet étant l'exposition de l'homme aux polluants du sol par le biais de la consommation de cultures, le chapitre 2 présente une synthèse bibliographique relative à l'assimilation des métaux par les plantes, et aux principaux modèles visant à estimer cette assimilation (et plus particulièrement aux limites de ces modèles). Le chapitre 3 présente deux méthodes pour prendre en compte l'incertitude liée aux paramètres des modèles d'exposition (les méthodes probabiliste et possibiliste) et propose une approche, dite « hybride », pour combiner ces deux méthodes dans une même évaluation du risque. Le chapitre 4 présente l'application de l'approche hybride à un cas réel de risque d'exposition à des métaux lourds du sol par le biais de cultures. Le chapitre 5 s'intéresse au problème dit « inverse » : compte tenu d'un risque d'exposition jugé « tolérable », et de l'incertitude liée aux paramètres d'exposition, quelle peut être la concentration maximale admissible dans le sol ? Le chapitre 6 présente des conclusions et des perspectives de développement futur.

2. Synthèse sur les modèles d'assimilation des métaux par les plantes

2.1. INTRODUCTION

Toutes les plantes ont la faculté d'accumuler, à partir du sol et de l'eau, les métaux qui sont nécessaires à leur croissance : Fe, Mn, Cu, Zn, Mg, Mo. Mais certaines plantes peuvent également accumuler des métaux lourds qui n'ont aucune fonction biologique connue : Cd, Cr, Pb, Co, Ag, Se, Hg. La capacité des plantes à assimiler des métaux toxiques contenus dans le sol, ou à réduire leur mobilité, connaît un intérêt considérable depuis une dizaine d'années. En effet, la phytoremédiation est de plus en plus souvent considérée comme une solution alternative pour le traitement des sites pollués par des éléments métalliques. Le terme générique de phytoremédiation désigne en fait plusieurs procédés (Schnoor, 1997) :

- la phytoextraction, au cours de laquelle les métaux contenus dans les sols sont extraits et concentrés par la plante dans sa partie moissonnée (Kumar et al., 1995),
- la rhizofiltration, au cours de laquelle des métaux toxiques contenus dans des effluents sont absorbés, précipités, ou concentrés par les racines des plantes,
- la phytostabilisation, au cours de laquelle des plantes qui ont une tolérance vis à vis des métaux lourds, réduisent leur mobilité en diminuant l'érosion et l'envol de poussières.

Le processus qui nous intéresse plus particulièrement dans le cadre de la présente étude est la phytoextraction, puisque la partie moissonnée de la plante est celle qui est susceptible d'être consommée par l'homme.

La capacité des plantes à extraire les métaux du sol dépend notamment de la disponibilité de ces métaux. Or ils peuvent être présents dans le sol sous des formes très diverses :

1) ions et complexes métalliques en solution dans l'eau interstitielle,
2) ions métalliques présents sur les sites d'échange des constituants inorganiques du sol,
3) métaux adsorbés à de la matière organique,
4) composés métalliques précipités ou insolubles (oxydes, hydroxydes, carbonates, sulfures), etc.

Une des difficultés liées à l'estimation du transfert sol-plante a trait au fait que seule la fraction 1), et probablement une partie de la fraction 2), sont aisément extraites par la plante. Pour que l'extraction par la plante ait lieu, il est nécessaire que le métal passe en solution. Cette mobilisation peut être effectuée par plusieurs mécanismes (Figure 1 A) :

1) des molécules (des phytosidérophores), qui chélatent les métaux (les intègrent dans leur structure), peuvent être sécrétées dans la rhizosphère pour solubiliser des métaux fixés dans le sol (Mench et Fargues, 1994),
2) les racines peuvent réduire des métaux fixés dans le sol, ce qui augmente leur mobilité,
3) les racines peuvent émettre des protons et acidifier le sol qui les entoure, ce qui a pour effet de solubiliser les métaux lourds.

Les métaux en solution peuvent pénétrer la racine par voie extracellulaire (apoplasmique ; Figure 1 A et Figure 2), intracellulaire (symplasmique ; idem), ou transcellulaire (Figure 2). La voie apoplasmique est la plus accessible à l'eau et aux ions minéraux (Heller et al., 1998). La plupart des ions métalliques pénètrent la plante par l'intermédiaire de porteurs ou de chenaux génériques (Clarkson et Luttge, 1989). La faible sélectivité des porteurs pour les métaux essentiels et non-essentiels expliquerait pourquoi ces derniers peuvent pénétrer la plante (même en présence d'un gradient de concentration contraire).

Une fois que les ions métalliques ont pénétré la racine ils peuvent être soit stockés, soit transportés vers la tige. Le transport vers la tige est lié essentiellement à l'appel créé par les parties aériennes (la transpiration). Des études (Salt et al., 1995) suggèrent que le transport de certains métaux non-essentiels vers la tige serait facilité par des acides organiques. La plupart des études témoignent de concentrations en métaux lourds non-essentiels nettement supérieures dans les racines que dans les tiges (Wheeler et Power, 1995). D'après Chaney et al. (1985), les concentrations en Zn, Cd, ou Ni dans les racines sont généralement au moins 10 fois supérieures à celles dans les tiges et les feuilles des plantes. Par contre, dans certaines plantes dites « hyperaccumulatrices », les concentrations dans les parties supérieures peuvent dépasser celles dans les racines (Brown et al., 1995).

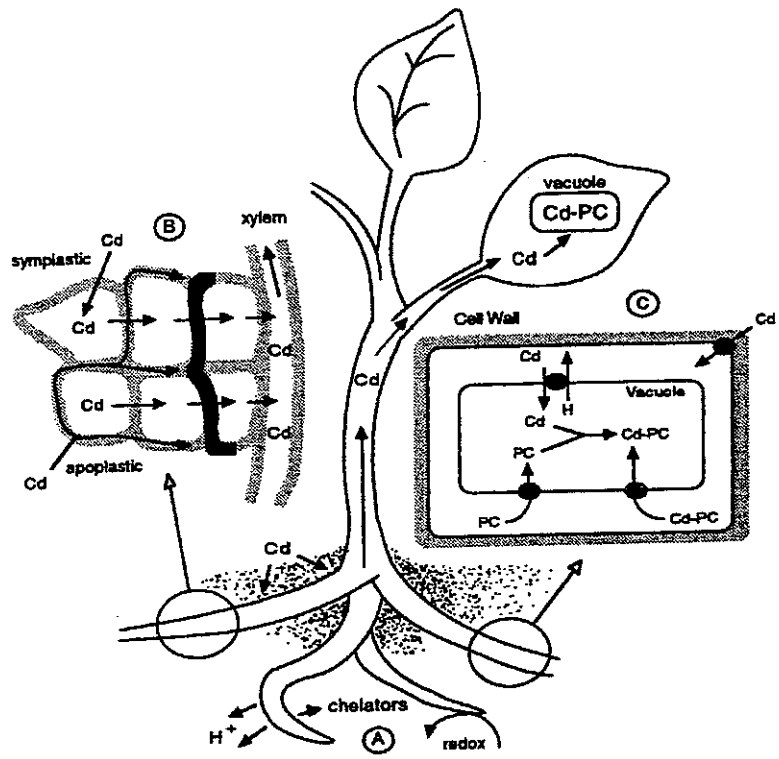


Fig.1 - Chemins possibles pour l'accumulation du Cd dans les plantes (Salt et al., 1995)

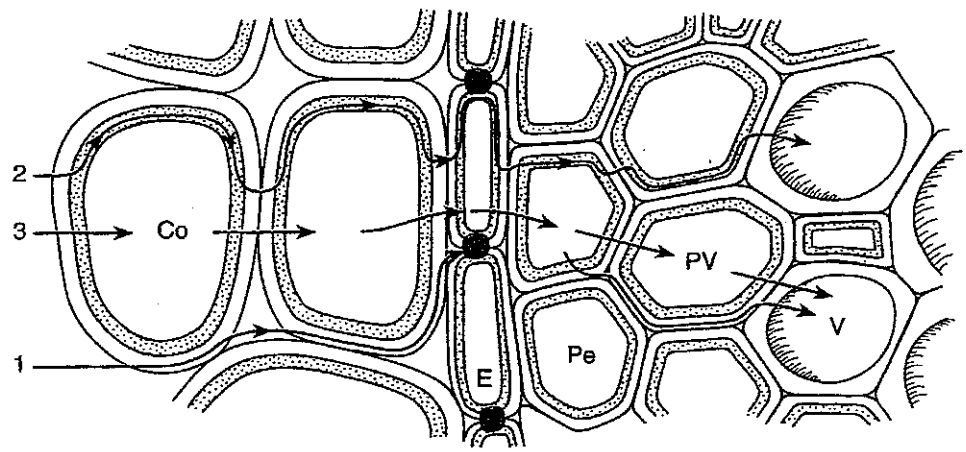


Fig.2 – Entrées possibles pour l'eau dans la racine (Heller et al., 1998)

1) apoplasmique ; 2) symplasmique ; 3) transcellulaire. Co : cellule du cortex ; E : endoderme ; Pe : péricycle ; PV : parenchyme vasculaire ; V : vaisseau.

2.2. LES FACTEURS QUI PEUVENT INFLUENCER LE TRANSFERT SOL-PLANTE

Comme il a été dit précédemment, un des principaux facteurs qui influence l'absorption des métaux par les plantes est la forme sous laquelle se trouvent les métaux dans le sol (ions libres, adsorbés, etc.). Les autres facteurs importants peuvent être subdivisés en facteurs qui sont liés à la plante, et ceux qui sont liés au sol. Certaines espèces de plantes ont la capacité d'émettre des agents chélatants qui permettent de solubiliser les métaux du sol. Mais deux éléments perturbateurs viennent compliquer les choses. D'une part, la variété du génotype au sein d'une même espèce aurait une influence importante sur la capacité des plantes à extraire les métaux (Michel Mench, comm. pers.). D'autre part, pour une même variété, l'influence de la série de sol (ludisols, calcosols, etc.) sur la phytodisponibilité des éléments métalliques serait prépondérante. Ces différents facteurs rendent particulièrement difficile l'élaboration d'un modèle fiable de transfert sol-plante.

Parmi les paramètres liés au sol on citera le pH du sol, qui est l'un des principaux paramètres influençant la mobilité des métaux. Le pH des sols varie généralement entre 4 et 8.5, avec un effet tampon par l'aluminium à faible pH et par le carbonate de calcium à pH élevé (Wild, 1988). La plupart des métaux lourds toxiques ont une mobilité accrue à faible pH. De nombreux auteurs constatent une nette augmentation de la concentration de métaux lourds dans les plantes lorsque le pH diminue, ou à l'inverse, une diminution de l'absorption par les plantes lorsque le pH augmente (Dijkshoorn et al., 1983; Singh et al., 1995 ; Xiong et Lu, 1993). Pour Dijkshoorn et al. (1983), l'importance de l'augmentation de l'absorption avec la diminution du pH diminuerait suivant la série Mn, Ni, Cd, Zn, Pb, Cu, Cr (voir aussi Smith, 1993). Ces auteurs notent que l'application de fertilisants a pour effet de diminuer les concentrations dans les plantes, l'utilisation des nitrates par les plantes ayant pour effet d'augmenter le pH du sol. Jopony et Young (1993) ont étudié la phyto-disponibilité du plomb dans des sols contaminés par des rejets miniers. Ils observent peu de corrélation entre l'extraction par les plantes et le pH ou la concentration totale en Pb dans le sol. Par contre, ces auteurs observent une bonne corrélation, dans une large gamme de pH, entre l'extraction par la plante et les concentrations en Pb dans des solutions équilibrées avec des suspensions de sol. Une faible corrélation entre l'absorption de Cd et Zn par les plantes et les concentrations totales dans les sols, est également constatée par Lorenz et al. (1997). Une corrélation bien plus significative est obtenue entre l'absorption et les concentrations dans l'eau interstitielle.

Le potentiel d'oxydo-réduction du sol a une influence importante sur la mobilité des métaux. Les conditions rédox du sol sont liées notamment à sa texture et à d'éventuelles pratiques agricoles. Des sols aérés auront un potentiel rédox élevé et la précipitation de métaux tels que le fer ou le manganèse sous forme d'oxydes sera favorisée. Pour le chrome au contraire, des conditions réductrices favoriseront la réduction du chrome VI, mobile et toxique, en chrome III, peu mobile et non toxique.

L'humidité et la température du sol favorisent, d'une part, la croissance de la plante, et d'autre part l'activité biologique dans la rhizosphère. Cette activité entraîne la production de molécules complexantes issues de la dégradation de substances organiques. L'association de ces molécules avec des métaux fixés dans le sol peut augmenter leur degré de mise en solution et donc leur phyto-disponibilité (Del Castilho et Chardon, 1995). D'après Chang et al. (1987), la température du sol serait l'un des principaux facteurs de variation influençant l'extraction des métaux par les plantes.

La compétition entre différents métaux lourds a une influence importante sur leur absorption par les plantes. Smilde et al. (1992) ont étudié l'absorption de Cd et Zn par différents végétaux élevés dans des pots (de la laitue, des endives, du maïs, du blé, des épinards), pour deux sols (un sol sableux et un sol argileux). Leurs résultats montrent que l'application de Cd au sol sableux contenant du Zn augmente l'extraction de Zn par la plante. A l'inverse, l'application de Zn à un sol sableux contenant du Cd diminue l'extraction de Cd par la plante. Ce dernier résultat est également obtenu pour le sol argileux. Par contre, l'application de Cd au sol argileux contenant du Zn diminue l'extraction de Zn par la plante (résultat inverse de celui obtenu pour le sol sableux). Ces travaux illustrent l'influence de la compétition entre les ions métalliques, et l'influence du type de sol. Un autre facteur d'influence mis en évidence par les auteurs est l'effet négatif de l'application de Zn sur la croissance de la plante. Pour Chaney et al. (1999), Zn inhibe à la fois l'absorption de Cd par la plante, et le transfert de Cd depuis les racines de la plante vers les parties aériennes.

La présence d'autres ions tels que Ca et Mn peut également influencer l'extraction de Cd par les plantes. Jarvis et al. (1976) montrent que la présence de ces ions en solution diminuent l'absorption de Cd par une herbacée, par un effet de compétition pour les sites d'échange à la surface des racines. La compétition est particulièrement importante entre Ca et Cd car ces ions ont des rayons très proches.

2.3. LES MODELES DE TRANSFERT SOL-PLANTE

2.3.1. Introduction

Au vu de la complexité des mécanismes impliqués dans la phyto-extraction des métaux lourds toxiques, il n'est pas surprenant qu'il existe à l'heure actuelle peu de consensus scientifique relatif à la manière adéquate de prévoir la concentration dans la plante à partir de la concentration dans le sol. Les modèles de calcul utilisés dans les outils d'évaluation des risques existants adoptent des hypothèses très simplificatrices (voire simplistes), en invoquant néanmoins le caractère « raisonnablement majorant » des résultats. Mais comme le montrent les calculs présentés plus bas, ce caractère « raisonnablement majorant » ne peut être supposé a priori.

2.3.2. Modèle A : HESP

Le modèle HESP (Human Exposure to Soil Pollutants ; Poels et al., 1990 ; Veerkamp et Berge, 1992) est un code de calcul de l'exposition de l'homme à des polluants du sol. Ce modèle considère de nombreuses voies d'exposition parmi lesquelles la consommation d'eau souterraine polluée, l'absorption de poussières, la consommation de végétaux, de volailles, etc. Ce modèle est utilisé surtout en Europe pour estimer les risques générés par les sites d'anciennes activités industrielles. Un autre modèle (CSOIL) est basé sur des hypothèses de calcul très semblables voire identiques.

Pour ce qui concerne l'exposition à des éléments métalliques par le biais des cultures, le modèle HESP cite les travaux de Baes (1982), qui utilise un facteur de bio-concentration (BCF) sol-plante :

$$BCF = \frac{C_{pl}}{C_s} \qquad (1)$$

où : C_{pl} est la concentration en métal dans la plante (masse de polluant par unité de masse de plante sèche) ; C_s est la concentration en métal dans le sol (masse de polluant par unité de masse de sol sec).

En faisant l'hypothèse que la phytodisponibilité est inversement corrélée au coefficient de partage solide-liquide (K_d), Baes propose une équation reliant ces deux variables, dont les coefficients ont été ajustés à partir de données présentées par l'auteur :

$$\ln[BCF)] = a - b\ln[Kd)] \qquad (2)$$

où : a et b sont des valeurs d'ajustement de cette équation aux valeurs observées utilisées par l'auteur (dans HESP on a : $a = 2.67$ et $b = 1.12$).

Les auteurs de HESP proposent des équations de corrélation entre le K_d pour différents métaux, et les proportions de matière organique et d'argile du sol. Pour Cd par exemple, l'équation est :

$$K_d = 95 + 1.7(A + 3MO) \qquad (3)$$

où : K_d est en l/kg ; A est la proportion d'argile du sol (en %) ; MO est la proportion de matière organique du sol (en %).

Une correction est effectuée pour prendre en compte une éventuelle différence entre le pH du sol sur lequel est cultivée la plante, et celui du sol sur lequel a été, le cas échéant, mesuré le K_d :

$$BCF^* = BCF\ x10^{-0.25(pHsol - pHKd)} \qquad (4)$$

où : BCF^* est le facteur de bio-concentration corrigé ; pH-sol est le pH du sol sur lequel est cultivée la plante ; pH-Kd est le pH de l'échantillon de sol sur lequel a été mesuré le K_d.

On remarque, par rapport à ce qui a été dit précédemment, qu'une lacune importante de ce modèle est qu'il se réfère à des concentrations totales en éléments métalliques dans le sol. Par ailleurs, la concentration dans la solution est abordée par l'intermédiaire du K_d, dont on sait pourtant qu'il est un descripteur peu prédictif du partage des métaux lourds entre la phase solide et liquide (large gamme de variabilité). Le K_d est utilisé pour décrire le partage entre une fraction de métal en solution et une fraction adsorbée. Or une concentration totale en métal dans le sol inclura des phases porteuses autres que la fraction adsorbée (carbonates, sulfures, phases silicatées, etc.).

Une prévision plus fiable de l'absorption par la plante serait obtenue si l'on connaissait la concentration en métal directement dans l'eau interstitielle du sol dans le sol superficiel. Ceci suppose des extractions de la solution du sol, extractions qui ne sont généralement pas réalisées lors des études de diagnostic des sites contaminés par des métaux. La littérature scientifique fait état de nombreuses expériences d'extraction à l'aide de différents extractants (solutions de sels neutres, DTPA, EDTA, NH_4NO_3, etc. ; voir par exemple Tessier et al., 1979). A noter que Hooda et Alloway (1993) observent une absence de corrélation entre la concentration en Cd et Pb extraite du sol par le DTPA, et l'absorption de métal par les plantes (une herbacée). Jopony et Young (1993) observent par contre une bonne corrélation entre l'absorption de Pb par les plantes et la concentration en Pb dans des solutions équilibrées avec des suspensions de sol. Les temps d'équilibrage utilisés sont de 14 jours (voir Jopony, 1991). A noter que le caractère minorant, par rapport à la biodisponibilité, des extraits à l'eau, a été mis en évidence notamment par Kahru et al. (2002), à partir d'essais comparatifs d'écotoxicité sur des extraits à l'eau et des suspensions eau-sol.

Une illustration des résultats obtenus à l'aide du modèle de HESP est présentée ci-dessous. On considère un sol dont la concentration en Cd est de 10 mg par kg de sol sec. On suppose que ce Cd est phyto-disponible. La proportion d'argile est de 10% et la proportion de matière organique de 5%. L'équation (3) donne K_d = 138 l/kg. La Figure 3 illustre l'influence du pH sur la concentration dans la plante, pour différentes valeurs de pH-Kd. Comme l'illustre cette figure, ce modèle tente de reproduire les observations reportées dans la littérature, qui témoignent d'une diminution de l'extraction par la plante lorsque le pH augmente. La correction introduite pour tenir compte de la différence entre le pH de l'échantillon sur lequel est mesuré le K_d (pH-Kd) est d'autant plus forte que pH-Kd est élevé. Bien sûr, lorsque pH-K_d = pH sol, chaque courbe donne la même valeur de concentration dans la plante. On note que sur cette gamme de valeurs de pH-Kd, les valeurs de C_{plante} calculées par HESP varient selon un facteur 3. La Figure 4 montre, pour un même pH du sol (pH-sol = 6), la sensibilité du calcul à la valeur de K_d. Compte tenu de l'incertitude relative à ce paramètre dans la pratique (au moins un ordre de grandeur), en particulier s'il est déterminé à partir de l'équation (3), cette figure illustre l'incertitude du résultat du calcul.

Une critique importante du modèle utilisé par le modèle HESP tel qu'il apparaît dans le manuel d'utilisation est qu'il ne tient pas compte des caractéristiques de la plante. La plante est considérée comme étant « générique » et le résultat du calcul sera a priori le même que l'on considère du tournesol ou du radis. C'est là une hypothèse extrêmement

simplificatrice auquel le modèle considéré plus bas tente de pallier. A noter, toutefois, que le rapport ECETOC (Poels et al., 1990), qui est à l'origine du modèle HESP, propose d'utiliser directement des valeurs de BCF sous forme de fourchettes, et présente des valeurs de BCF pour différents types de plantes (valeurs de BCF basses ou hautes).

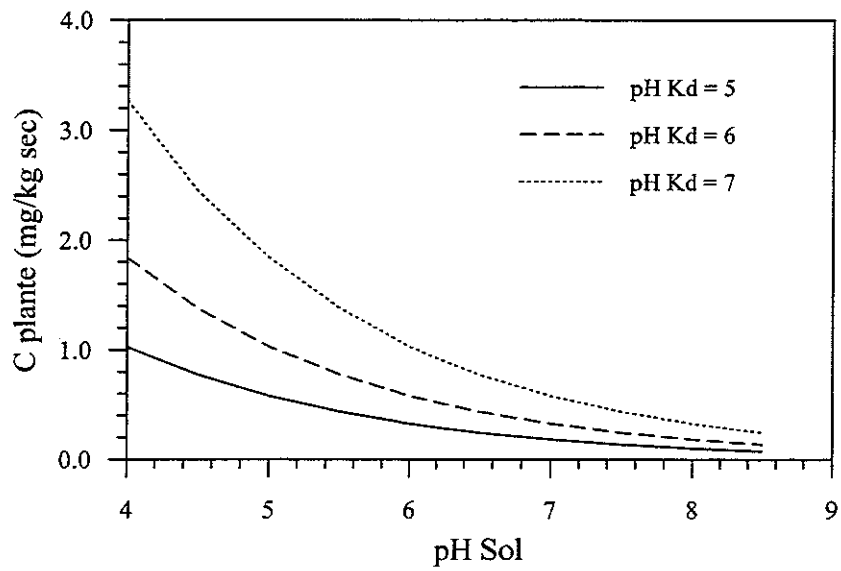


Fig.3 – Illustration de l'influence du pH sur la quantité de Cd extraite calculée par le modèle HESP

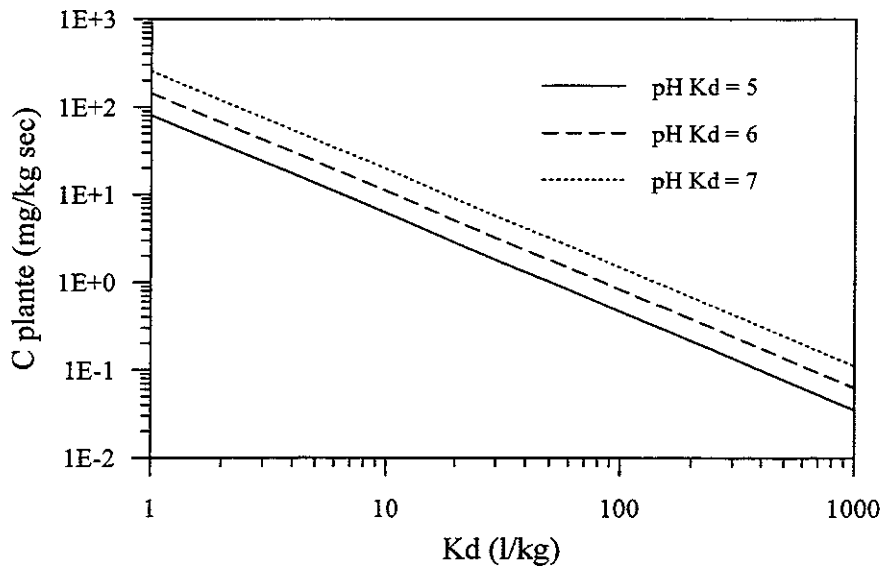


Fig.4 – Illustration de l'influence du K_d sur la quantité de Cd extraite calculée par le modèle HESP

2.3.3. Modèle B

Ce modèle combine une équation de Del Castilho et Chardon (1995), qui tente de corréler la concentration dans le sol avec celle en solution, avec une équation simple d'absorption par consommation racinaire. L'équation de consommation racinaire est (Palm, 1994) :

$$Ex_{plante} = C_{solution} \times r \qquad (5)$$

où : Ex_{plante} est la quantité de métal qui est extraite par la plante (mg/jour et par m^2 de culture) ; $C_{solution}$ est la concentration en métal dans l'eau du sol dans la zone racinaire (mg/l) ; r est la consommation d'eau racinaire (mm/j).

Del Castilho et Chardon (1995) ont étudié l'absorption de Cd par différents végétaux. Ces auteurs supposent qu'à de faibles concentrations en Cd dans le sol, la distribution du Cd entre la solution et le sol peut être décrite par une isotherme de Freundlich :

$$Cd_{ads} = K\, Cd_{solution}^{1/n} \qquad (6)$$

où : Cd_{ads} est la fraction adsorbée (mg/kg) ; $Cd_{solution}$ est la concentration en solution (µg/l) ; n et K sont des coefficients spécifiques au sol.

En se basant sur l'étude de 11 sols néerlandais, ces auteurs proposent une valeur de 1/n = 0.8 et une corrélation entre le paramètre K et le pH du sol :

$$K = -0.328 + 0.053 \times 10^{0.26\text{pH-sol}} \qquad (7)$$

Ce modèle est utilisé ci-dessous pour estimer la concentration de Cd dans la solution à partir de la concentration en Cd dans le sol. Cette concentration en solution sert ensuite pour estimer la concentration en Cd dans la plante en fonction d'une consommation racinaire (équation 5).

On considère, comme précédemment, un sol dont la concentration en Cd est de 10 mg/kg. Ce sol est cultivé par du blé dont la période de croissance est supposée de 150 jours et le rendement est de 0.3 kg/m^2 (Veerkamp et Berge, 1992). La consommation racinaire est calculée sur la base des valeurs d'évapotranspiration journalières (Penman-Monteith) mesurées sur un site près du Creusot (71) sur une période d'un an et demie, et en appliquant un bilan de Thornthwaite (Thornthwaite et Mather, 1955) pour estimer l'évapotranpiration « réelle » (ETR). On suppose une réserve utile (RFU) du sol de 50 mm et un coefficient cultural de 1 pour la période de février à juin (voir Nouzille et al., 1998). Une valeur d'ETR moyenne sur la période considérée de 2.3 mm/j est obtenue.

La Figure 5 présente les valeurs de concentration en Cd dans la plante estimées à l'aide du modèle B. On remarque des concentrations très fortes à faible pH. Ceci est lié au fait, d'une part, qu'à faible pH les équations (6) et (7) prévoient une concentration dans le fluide interstitiel relativement élevé, et d'autre part, il est supposé que le sol peut

« fournir » cette concentration tout au long de la croissance de la plante. Dans la réalité on peut s'attendre à une déplétion de la quantité de Cd dans le sol, ainsi qu'à un effet de la concentration en Cd de la plante sur sa croissance et donc sur sa consommation racinaire.

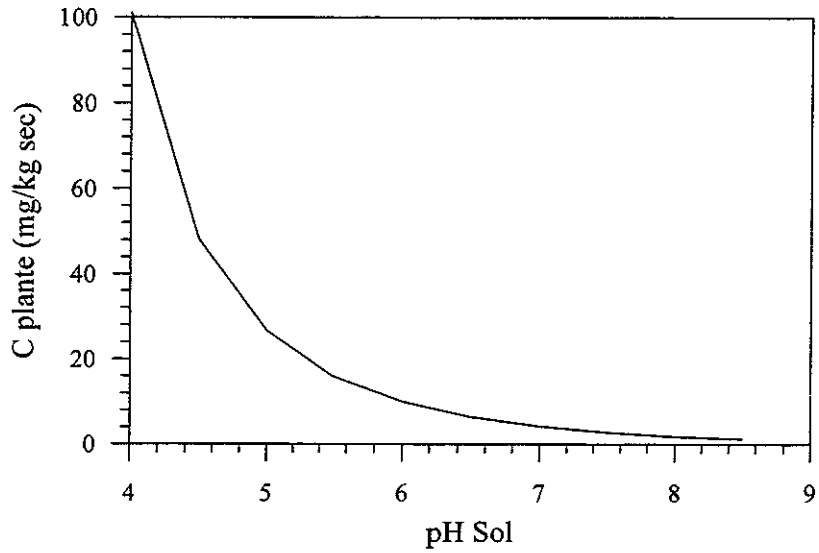


Fig.5 – Illustration de l'influence du pH sur la quantité de Cd extraite calculée par le modèle B

2.3.4. Comparaison avec des données mesurées

Les deux modèles décrits précédemment sont comparés aux données mesurées par Singh et al. (1995). Ces auteurs ont étudié l'influence du pH sur l'extraction de Cd par différents végétaux cultivés dans deux types de sols : un limon argileux et un limon. Les sols étudiés ont les caractéristiques suivantes :

Sol	pH	Carbone total (g/kg)	%Argile	% Sable	% Silt	Cd (mg/kg)*
Limon argileux	6.5	64	28	45	41	2.8
Limon	6.5	37	15	31	40	1.76

(* : Extrait à l'eau régale)

Tableau 1 – Caractéristiques des sols étudiés par Singh et al. (1995)

Le calcul est effectué pour des cultures de blé et de carottes. Ces végétaux ont été cultivés dans des pots de 7.2 dm^3 au cours de 3 années consécutives. Les pH des sols ont été ajustés par ajout de H_2SO_4 et $CaCO_3$ pour arriver à 4 valeurs de pH : 5.5, 6.5, 7.0 et 7.5. Les végétaux ont été analysés à maturité et la concentration en Cd a été mesurée dans différentes parties de la plante. Sur les 3 années d'expérimentation, on obtient les valeurs moyennes suivantes :

	Limon argileux		Limon	
pH	Carottes	Blé	Carottes	Blé
5.5	3.65	1.70	5.08	2.03
6.5	1.41	0.61	1.90	0.77
7	1.52	0.81	1.67	0.78
7.5	0.92	0.65	1.28	0.74

Tableau 2 – quantités de Cd (mg par kg de végétal sec) mesurées dans les végétaux.

Ces valeurs sont comparées ci-dessous aux résultats calculés par le modèle HESP et le modèle B décrits précédemment.

L'approche HESP nécessite une valeur de K_d. A partir de la composition des sols et de l'équation (3) (utilisée par HESP) on arrive à des valeurs de K_d de 175 l/kg pour le limon argileux et 139 l/kg pour le limon. Il est supposé que ce K_d correspond à un pH qui est celui du sol avant ajouts effectués pour modifier le pH (pH = 6.5).

Pour appliquer le modèle B, il a été supposé une durée de croissance des carottes et du blé de 90 jours (Singh et al., 1995) et 150 jours respectivement. La consommation racinaire a été prise égale à 2.3 mm/j (voir sous-chapitre précédent). Les rendements des deux cultures sont prises égales à 0.38 et 0.30 kg/m2 respectivement (Veerkamp et Berge, 1992).

Les résultats des calculs sont présentés dans les tableaux 3 et 4 et graphiquement dans les Figures 6 et 7. On constate que les valeurs calculées par le modèle HESP sont très fortement sous-estimées par rapport aux mesures. Les valeurs calculées par HESP sous-estiment les valeurs mesurées d'un facteur 11 à 28 pour les carottes, et d'un facteur 5 à 13 pour le blé. Le modèle B présente également des écarts importants par rapport aux valeurs mesurées, avec une sous-estimation d'un facteur 2 à 7 pour les carottes, et une sur-estimation jusqu'à un facteur 3 pour le blé.

pH	HESP	Modèle B	
		Carottes	Blé
5.5	0.22	1.77	3.74
6.5	0.12	0.72	1.51
7	0.09	0.47	0.99
7.5	0.07	0.31	0.66

Tableau 3 – valeurs de Cd (mg par kg de végétal sec) calculées par les deux modèles pour le limon argileux.

pH	HESP	Modèle B	
		Carottes	Blé
5.5	0.18	0.99	2.09
6.5	0.10	0.40	0.85
7	0.08	0.26	0.56
7.5	0.06	0.18	0.37

Tableau 4 – valeurs de Cd (mg par kg de végétal sec) calculées par les deux modèles pour le limon.

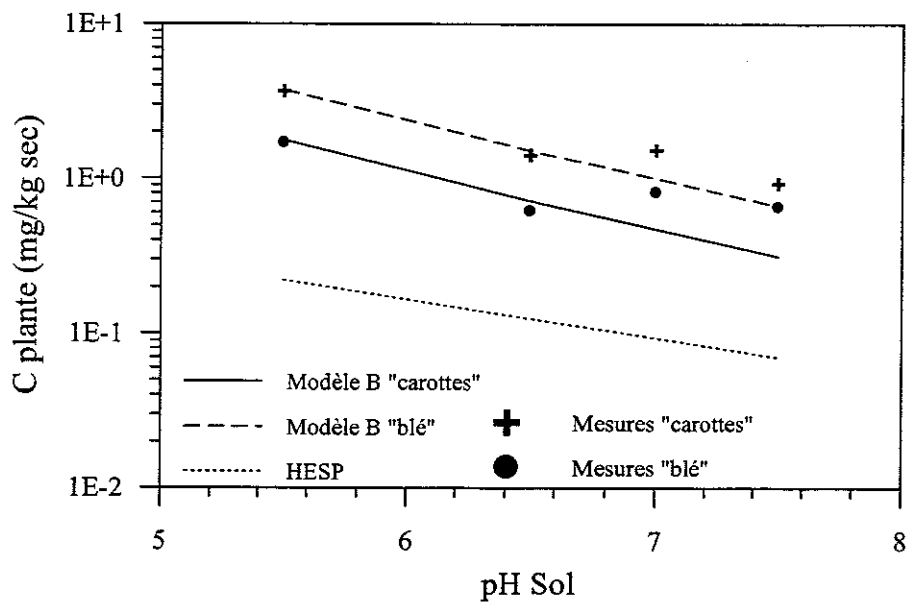


Fig. 6 – Comparaison entre les valeurs mesurées par Singh et al. (1995) et les valeurs calculées à l'aide de HESP et du modèle B. Sol : limon argileux

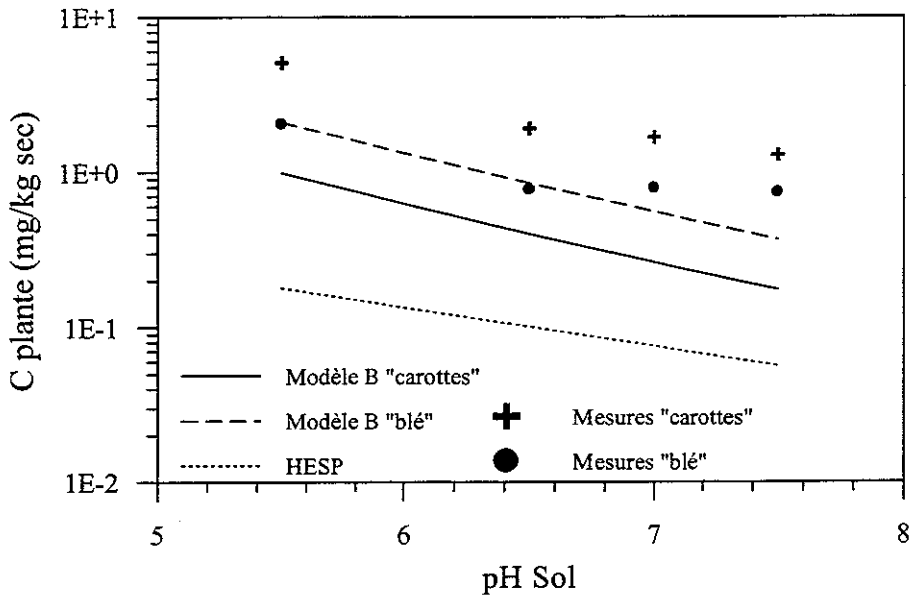


Fig. 7 – Comparaison entre les valeurs mesurées par Singh et al. (1995) et les valeurs calculées à l'aide de HESP et du modèle B. Sol : limon

2.3.5. Conclusions sur les modèles d'assimilation des métaux par les plantes

La revue bibliographique succincte présentée ci-dessus souligne les limites des modèles communément utilisés par les outils d'évaluation des risques pour estimer le transfert sol-plante. Compte tenu de la complexité des mécanismes en jeu, aucun des deux modèles testés ne peut être jugé satisfaisant. D'après la littérature, il y aurait peu de corrélation entre la concentration totale dans le sol et ce qui est réellement extrait par la plante. Par contre, certaines études montrent une corrélation significative entre la quantité de métal extraite et la concentration dans l'eau interstitielle du sol. De nombreux extractants (solutions de sels neutres, DTPA, EDTA, NH_4NO_3, etc.) ont été testés par les chercheurs pour accéder à cette solution interstitielle.

Compte tenu de ce qui précède, il paraît clairement préférable de disposer directement de mesures de l'assimilation de métaux par des plantes cultivées sur le site même. Or c'est un des principaux avantages du jeu de données analysé dans le chapitre 4 que d'inclure des mesures d'assimilation de métaux par des légumes cultivés sur le site industriel étudié.

3. Prise en compte de l'incertitude dans l'évaluation du risque d'exposition

3.1. INTRODUCTION

Comme il a été souligné dans l'introduction générale, le principal objectif de l'étude proposée est de permettre de combiner plusieurs modes de représentation de l'incertitude dans un même calcul du risque d'exposition. Dans un premier temps, les méthodes probabiliste et possibiliste de représentation de l'incertitude sont décrites, et ensuite une approche (appelée « hybride ») pour combiner ces deux méthodes est proposée.

3.2. REPRESENTATIONS PROBABILISTE ET POSSIBILISTE DE L'INCERTITUDE

Dans le cadre des évaluations des risques d'exposition de l'homme aux polluants du sol, l'approche probabiliste est couramment utilisée pour représenter l'incertitude liée aux paramètres des modèles d'exposition. L'approche probabiliste suppose que les paramètres des modèles d'exposition sont des variables aléatoires pouvant être décrites par des fonctions de distribution de probabilité. Si on considère une variable aléatoire X dont les probabilités P(x) associées à chaque valeur x de la variable X sont connues, la fonction de distribution $F(x)$ (notée aussi fonction de distribution cumulée) décrit la probabilité que la variable X prenne une valeur inférieure ou égale à une valeur x :

$$F(x) = P(X \leq x) \qquad \text{pour tout } x \tag{8}$$

Si la variable X est continue, alors on peut définir une fonction de densité de probabilité $f(x)$:

$$f(x) = \frac{dF(x)}{dx} \tag{9}$$

La distribution normale est un exemple bien connu de la fonction de densité de probabilité. La Figure 8 illustre cette fonction ainsi que la fonction de distribution de probabilité cumulée $F(x)$.

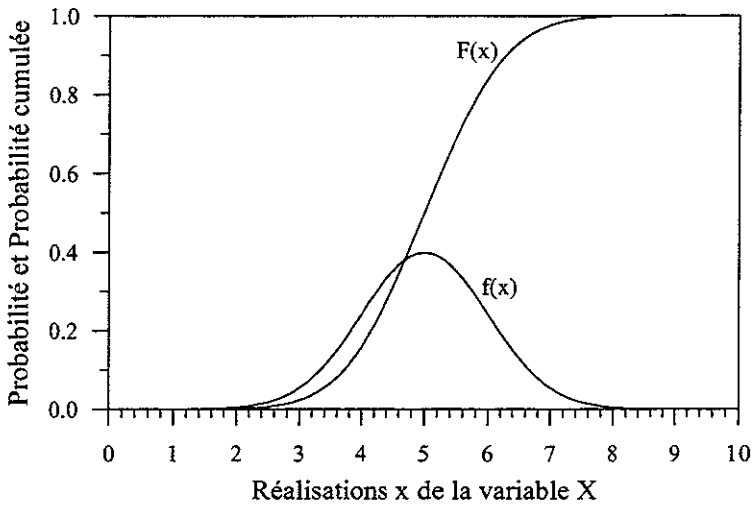


Fig. 8 - Exemple de fonction normale de densité de probabilité f(x) et fonction associée de distribution de probabilité cumulée F(x).

L'utilisation de l'approche probabiliste pour représenter l'incertitude suppose qu'on dispose de suffisamment de données pour définir des distributions de probabilité statistiquement significatives. Or dans la pratique, une telle quantité d'information fait souvent défaut. L'information dont on dispose est souvent de nature imprécise ou vague.

Certains chercheurs ont proposé de supposer « a priori » des distributions de probabilité pour représenter l'incertitude, même en l'absence d'informations permettant d'effectuer une analyse statistique. C'est l'approche dite « Bayésienne » des probabilités subjectives (voir par exemple Cullen et Frey, 1999) appliquée notamment par Bonano et al. (1990) dans le domaine de la sûreté nucléaire. Ces auteurs font appel à du jugement d'expert pour identifier des distributions de probabilité jugées les plus vraisemblables.

Dans le contexte de l'évaluation des risques pour la santé humaine, l'approche Bayésienne ne nous paraît pas satisfaisante pour plusieurs raisons. Tout d'abord elle suppose plus de connaissance qu'il n'en existe réellement : elle ne distingue pas entre le cas où on ne dispose pas d'information concernant un paramètre, de celui où on a réellement observé une variabilité aléatoire de ce paramètre. Ensuite, comme il sera montré dans le chapitre 4, l'hypothèse *a priori* de distributions de probabilité peut avoir des conséquences non sécuritaires (minorantes) sur l'estimation du risque. La vraie « force » de la théorie Bayésienne réside dans l'application du théorème de Bayes, qui permet de corriger les distributions de probabilité subjectives au fur et à mesure que de nouvelles informations sont recueillies. Cette théorie a été appliquée avec succès, notamment dans le domaine de l'exploration minière. Mais dans un contexte d'évaluation des risques, il est rare que de nouvelles informations soient disponibles permettant de corriger les distributions initiales. Ces distributions *a priori* deviennent donc *de facto* des distributions *a posteriori*, introduisant ainsi une illusion de connaissance (et de degré de confiance) qui n'est pas justifiée par les données disponibles.

Le besoin de traiter des données de type imprécises ou vagues a amené Zadeh (1965 et 1978) à développer une théorie dite des « possibilités », par opposition à la théorie plus connue des probabilités. Plutôt que de représenter une variable par une distribution de probabilité, Zadeh introduit la notion de nombre dit « flou ». Un tel nombre décrit, au travers d'une fonction d'appartenance (notée μ), le degré de « possibilité » qu'une variable X prenne telle ou telle valeur x. Ce degré de possibilité est analogue à la notion « d'absence de surprise ».

Pour illustrer le nombre flou on supposera que l'on dispose de quelques mesure du pH du sol, qui indiquent une valeur située entre 6.5 et 8 mais sans qu'il soit possible de définir une véritable distribution statistique pour ce paramère (faute d'avoir un nombre suffisant de mesures). Par contre, l'expérimentateur pourrait estimer que des valeurs minimales et maximales de 6 et 8.5 sont possibles, compte tenu du type de sol ou du contexte. Cette information peut être représentée par un nombre flou tel que décrit dans la Figure 9.

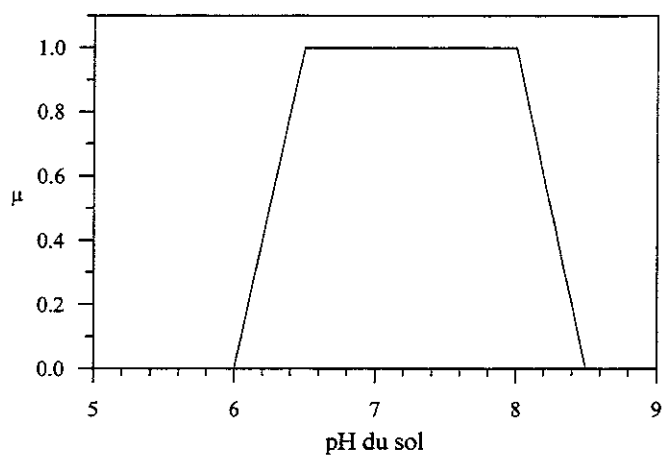


Fig. 9 - Exemple d'un nombre flou représentant les valeurs jugées vraisemblables du pH d'un sol

La fourchette 6.5 – 8 est appelée le « noyau » du nombre flou, tandis que la fourchette 6 – 8.5 est appelée le « support ». A noter que toutes les formes de nombres flous sont possibles, mais que leur définition devrait s'attacher à être cohérente avec l'information dont on dispose. Par exemple, si l'expérimentateur estime que les valeurs en dehors de la fourchette 6.5 – 8 sont possibles mais peu vraisemblables, les droites reliant le noyau au support pourraient être représentées par des courbes (Figure 10).

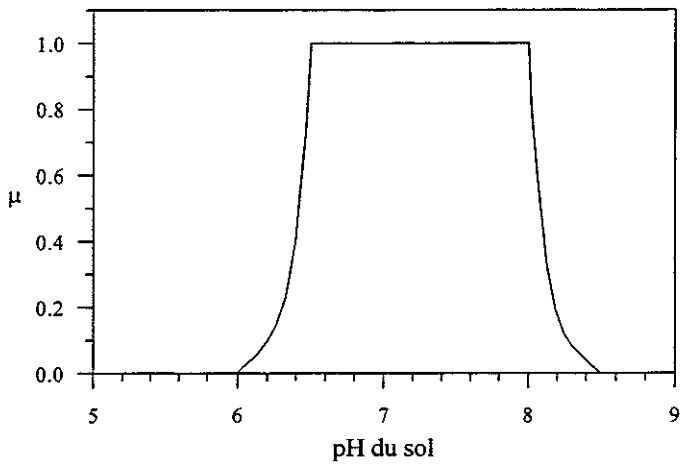


Fig. 10 - Même nombre flou qu'en Figure 9, mais les valeurs situées en dehors du noyau sont jugées moins vraisemblables.

Comme il sera montré dans le prochain sous-chapitre, la théorie des possibilités permet d'effectuer des calculs à partir de nombres flous (voir Prade et Dubois, 1988).

3.3. CALCUL MONTE CARLO ET CALCUL EN NOMBRES FLOUS

Si les variables intervenant dans un modèle sont représentées par des distributions de probabilité, un calcul probabiliste peut être effectué en utilisant la méthode dite « Monte Carlo » (voir Vose, 1996). Cette méthode consiste à échantillonner, de manière aléatoire, les différentes fonctions de distribution de probabilité (appelées PDF en anglais), à effectuer le calcul à partir du modèle, puis à répéter l'opération un grand nombre de fois. Dans la pratique, les fonctions de probabilité sont échantillonnées « au hasard » en générant des nombres aléatoires compris entre 0 et 1 (il existe pour cela des algorithmes spécifiques). Cette procédure est illustrée graphiquement dans la Figure 11. On voit que si le tirage aléatoire donne une valeur $\alpha = 0.65$, on obtient une valeur de la variable X : $x = 5.4$.

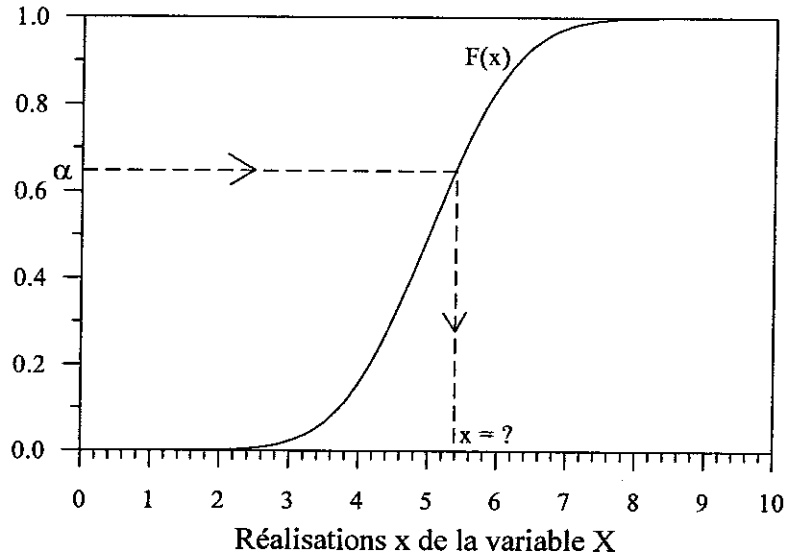


Fig. 11 - Principe de l'échantillonnage aléatoire des distributions de probabilité cumulée

Si les différentes variables X_1, X_2, .., X_k du modèle sont supposées indépendantes, un nombre aléatoire est tiré pour chaque variable et ensuite le calcul du modèle f(X_1, X_2, .., X_k) est effectué. En répétant la procédure un grand nombre de fois (on parle d'itérations), on peut exprimer le résultat du calcul sous la forme d'un histogramme de fréquences relatives, et d'un diagramme des fréquences relatives cumulées. Ce diagramme est analogue à une distribution de probabilité cumulée. Il peut ensuite être comparé à un critère d'acceptabilité (voir la section suivante).

Le calcul à partir de nombres flous, quant à lui, fait usage du principe d'extension énoncé par Zadeh (1978). Ce principe permet d'étendre aux fonctions d'appartenance, les opérations arithmétiques effectuées sur les valeurs d'un sous-ensemble flou. Il énonce qu'étant donné deux quantités floues X_1 et X_2 non liées, et une fonction réelle $f(x,y)$, la fonction d'appartenance de la quantité floue $f(X_1, X_2)$ est obtenue à partir des fonctions d'appartenance de X_1 et X_2 d'après :

$$\mu_{f(X_1,X_2)}(u) = Sup\left\{min\left(\mu_{X_1}(x), \mu_{X_2}(y)\right) \middle| f(x,y) = u\right\} \qquad (10)$$

où : $\mu_{f(X_1,X_2)}(u)$ est la fonction d'appartenance de $f(X_1,X_2)$ pour toute valeur u ; $\mu_{X_1}(x)$, $\mu_{X_2}(y)$ sont les fonctions d'appartenance des nombres flous X_1 et X_2 pour les valeurs x et y ; *Sup* désigne la plus grande valeur (le suprémum) ; et *min* désigne l'opérateur de minimisation des $\mu_{X_1}(x)$, $\mu_{X_2}(y)$ tels que $f(x,y) = u$.

L'équation (10) permet de définir les règles d'une arithmétique en nombres flous. On citera Bouchon-Meunier (1993) pour des applications de l'équation (10) à des opérations simples telles que l'addition, la soustraction ou la multiplication. Par contre, son application à des fonctions complexes peut se révéler très lourde. Une alternative pratique pour réaliser des opérations à partir de nombres flous est la méthode par découpage. Cette méthode consiste à sélectionner la gamme de valeurs (appelée α-

coupe) envisageables pour une certaine valeur α de la fonction d'appartenance d'un nombre flou X (voir Figure 12), et à faire usage de la relation ci-dessous.

Soit k quantités floues X_1 à X_k, et une fonction continue f de ces quantités (le modèle), on peut montrer que :

$$\left[f(X_1,...,X_k)\right]_\alpha = f(X_{1\alpha},...,X_{k\alpha}) \qquad (11)$$

pour toute valeur de α comprise entre 0 et 1.

Cette relation exprime le fait que la α-coupe correspondant au résultat du modèle est égale au résultat du modèle obtenu à partir des α-coupes de chaque quantité floue X_i intervenant dans le modèle. Dans la pratique, et pour le type de fonctions considérées ici, la méthode des α-coupes revient à effectuer un calcul d'intervalle classique pour différentes valeurs α de la fonction d'appartenance, et à reconstituer ainsi le résultat flou.

A noter qu'une différence fondamentale entre les approches floue et probabiliste est la non-multiplicité des degrés de vraisemblance. Il est rappelé, en effet, qu'étant donné deux événements A et B indépendants, la probabilité que A et B se réalisent simultanément est le produit des probabilités respectives de réalisation de A et de B :

$$P(A \cap B) = P(A) \cdot P(B) \qquad (12)$$

Lors de simulations Monte Carlo, les scénarios associant plusieurs événements peu probables ont d'autant moins de chances d'être réalisés. L'approche floue, par contre, s'intéressant aux possibilités et non aux probabilités, ne répercute pas les degrés de vraisemblance respectifs des différentes variables dans celle du scénario tout entier.

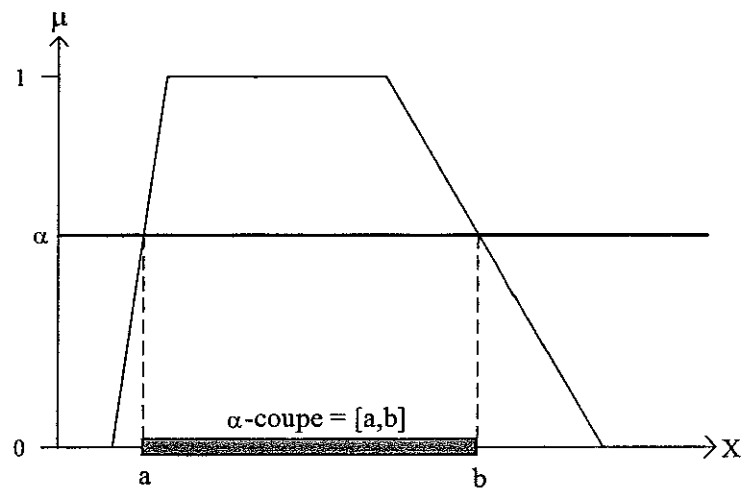


Fig. 12 - Principe de la méthode par découpage (α-coupes).

En se basant sur l'équation (11), le calcul flou peut être effectué en recherchant les minimas et les maximas du calcul du modèle, pour un nombre limité de niveaux de vraisemblance (valeurs α de la fonction d'appartenance), et en reconstituant le résultat flou complet en reliant les points ainsi définis. Pour k variables floues X_1 à X_k, le schéma de calcul est le suivant :

① : choisir une valeur α de la fonction d'appartenance (un niveau de vraisemblance),
② : sélectionner, pour chaque nombre flou X_1 à X_k, les valeurs a et b de la α-coupe correspondant à cette valeur α,
③ : calculer les valeurs *min* et *max* de $f(X_1, ..., X_k)$ en considérant toutes les valeurs situées entre a et b pour chaque nombre flou,
④ : affecter ces valeurs *min* et *max* aux limites inférieure et supérieure de la α-coupe de $f(X_1, ..., X_k)$,
⑤ : répéter l'opération pour une autre α-coupe,
⑥ : reconstituer le résultat flou $f(X_1, ..., X_k)$ à l'aide des valeurs *min* et *max* de chaque α -coupe de $f(X_1, ..., X_k)$.

Cette procédure a été appliquée par Guyonnet et al. (1999) pour estimer l'impact d'une source de pollution sur les eaux souterraines. Ces auteurs ont également comparé les méthodes probabiliste (Monte Carlo) et possibiliste (calcul en nombres flous) pour un même scénario. Il est montré combien l'approche floue est sécuritaire par rapport à l'approche Monte Carlo. En effet, l'approche par calcul flou considère toutes les combinaisons possibles de valeurs de paramètres mais ne multiplie pas les degrés de vraisemblance (ou d'invraisemblance). Dans le calcul Monte Carlo au contraire, des scénarios qui combinent des valeurs peu probables des variables ont une probabilité quasi-nulle. A supposer qu'on échantillonne ces scénarios lors du tirage aléatoire, ils seront éliminés lorsqu'on comparera le résultat du calcul avec un critère d'acceptabilité, pour un certain pourcentage de confiance (95% par exemple). De nombreux scénarios peu probables se trouveront dans cette gamme de 5% restants. Si les valeurs des paramètres avaient été jugées « possibles » plutôt que « probables » cet effet de coupure n'aurait pas lieu. Cet effet minorant de l'approche Monte Carlo est illustrée plus en détail dans le chapitre 4, par comparaison avec la méthode « hybride » proposée.

3.4. COMPARAISONS AVEC DES CRITERES DE TOLERANCE

La Figure 13 présente un exemple de résultat de calcul Monte Carlo du risque de cancer additionnel lié à l'ingestion d'eau souterraine polluée par un produit chimique (du trichloréthylène ; voir Guyonnet et al., 1999). Le risque calculé est comparé à un critère d'acceptabilité défini par l'autorité compétente comme étant 10^{-5}. Cette valeur signifie que l'on accepte que dans une population de 10^5 personnes, un cancer additionnel soit imputable à la présence de la source de pollution. L'autorité compétente stipule par ailleurs que le risque calculé ne doit pas dépasser le critère d'acceptabilité pour un degré de confiance de 70%. La Figure 13 indique que le risque calculé est jugé acceptable. En effet, on voit que le risque calculé a plus de 70% de chances d'être inférieur à 10^{-5}.

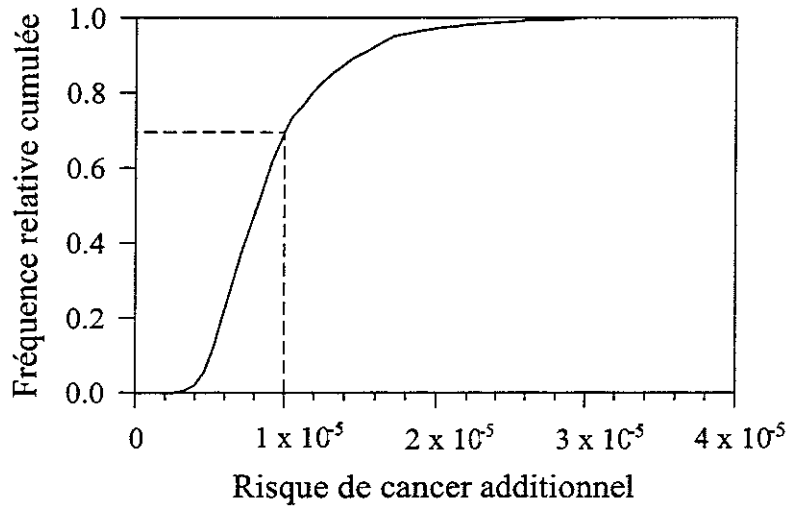


Fig. 13 - Comparaison du risque calculé (par la méthode Monte Carlo) avec un critère d'acceptabilité.

La comparaison d'un résultat flou avec un critère d'acceptabilité est légèrement moins « intuitive ». La théorie des possibilités permet la comparaison d'un critère (noté C) avec un nombre flou (noté F), en faisant intervenir deux indicateurs de la véracité de la proposition "F est inférieure à C". Ces indicateurs, compris entre 0 et 1, sont appelés mesure de possibilité (Π) et mesure de nécessité (N). Ils peuvent être définis comme suit (Prade et Dubois, 1988, Bouchon-Meunier, 1993).

La mesure de possibilité Π que la proposition $F \leq C$ soit vraie est définie par :

$$\Pi(F \leq C) = \underset{u}{Sup}\ min\left[\mu_F(u), \mu_C(u)\right] \qquad (13)$$

où : $\mu_F(u)$ est le degré d'appartenance de la grandeur F pour toute valeur u ; $\mu_C(u)$ est le degré d'appartenance du critère C pour toute valeur u (note : $\mu_C(u) = 1$ si $u \leq C$ et 0 si $u > C$) ; *Sup* désigne la plus grande valeur et *min* désigne l'opérateur de minimisation.

Dans l'équation (13), $min\left[\mu_F(u), \mu_C(u)\right]$ représente l'ensemble des valeurs correspondant aux plus petits des degrés d'appartenance de F et de C. La mesure de nécessité N est définie par :

$$N(F \leq C) = \underset{u}{Inf}\ max\left[1 - \mu_F(u), \mu_C(u)\right] \qquad (14)$$

où *Inf* désigne la plus petite valeur et *max* désigne l'opérateur de maximisation.

A partir de ces deux indicateurs, on peut distinguer 5 situations illustrées en Figure 14 pour un nombre flou F d'allure générale trapézoïdale, un critère C (supposé non-flou), et la proposition $F \leq C$:

- Figure 14 a) : $\Pi = 1$, $N = 1$. La proposition est nécessairement vraie,
- Figure 14 b) : $\Pi = 1$, $N = 1-\alpha$. La proposition est vraisemblable, pour une nécessité de $1-\alpha$,
- Figure 14 c) : $\Pi = 1$, $N = 0$. La proposition est peut-être vraie,
- Figure 14 d) : $\Pi = \beta$, $N = 0$. La proposition est vraisemblablement fausse, mais peut être vraie pour une possibilité de β,
- Figure 14 e) : $\Pi = 0$, $N = 0$. La proposition est nécessairement fausse.

On notera que N est toujours inférieur à Π, et que si $N = 1$, alors Π est forcément aussi égal à 1. En effet, pour une proposition A on a : $N(A) = 1 - \Pi(\overline{A})$, où $(\overline{A})$ représente la proposition contraire de A. Or si la mesure de nécessité de la proposition A est égale à 1, alors la mesure de possibilité de la proposition contraire de A est égale à 0. Par conséquent la mesure de possibilité de la proposition A est égale à 1.

Pour la proposition considérée dans la Figure 14, il est conseillé (voir Côme et al., 1997) de sélectionner N comme véritable indicateur de l'acceptabilité, car ce critère est un critère "strict", compatible avec une application raisonnable du principe de précaution.

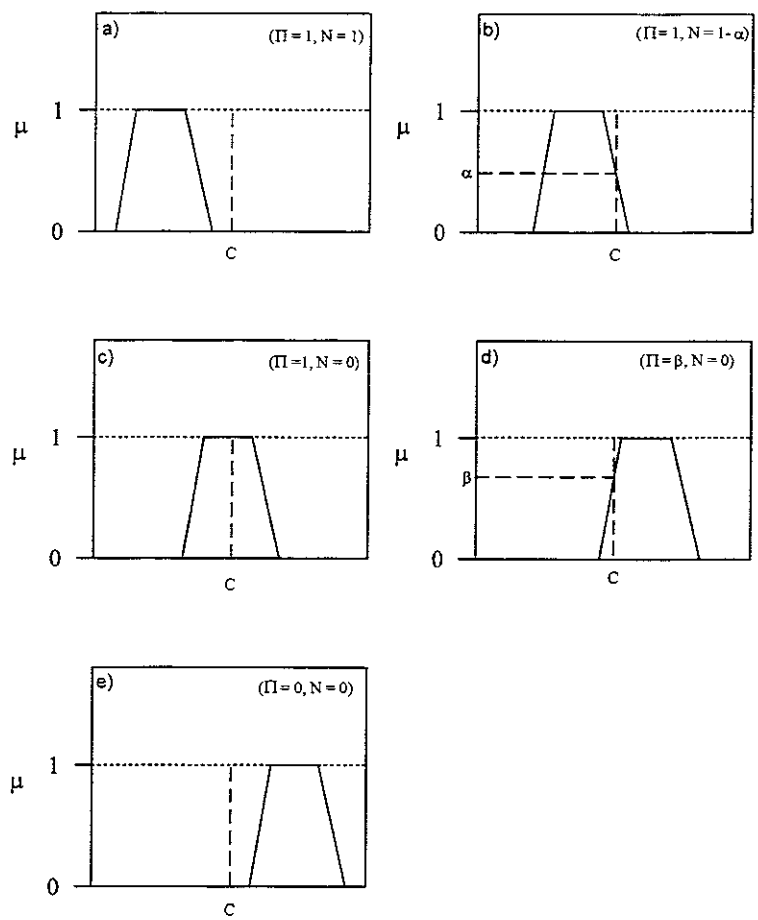


Fig. 14 - Comparaison d'un résultat flou (F) avec un critère d'acceptabilité (C) pour la proposition : résultat flou F est inférieur ou égal à C

La Figure 15 compare un résultat de risque flou à un seuil en utilisant les indicateurs présentés plus haut. En supposant que l'autorité compétente stipule que le risque calculé doive être inférieur à 10^{-5} pour une valeur de nécessité de 70 %, la Figure 15 suggère que le risque soit jugé inacceptable. En effet, la valeur de risque correspondant à cette valeur de nécessité est : 3.5×10^{-5}.

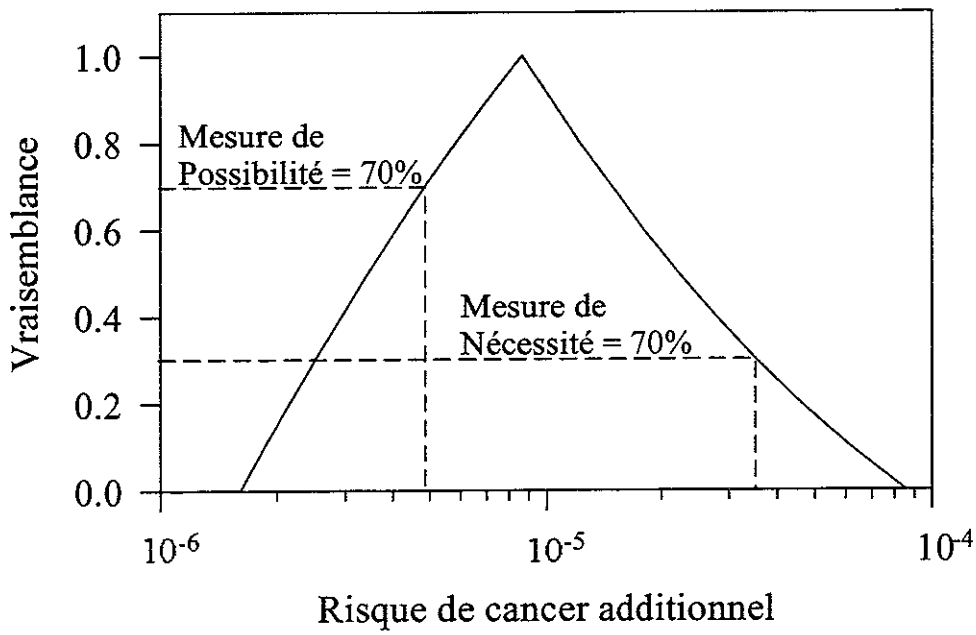


Fig. 15 - Valeurs maximales de risque calculé pour des mesures de possibilité (Π) et de nécessité (N) de 70%.

3.5. DEVELOPPEMENT D'UNE APPROCHE « HYBRIDE » PROBABILISTE-POSSIBILISTE

L'approche « hybride » proposée a été développée en collaboration avec l'Université Paul Sabatier de Toulouse. Elle permet de combiner, dans un même calcul du risque, la méthode dite Monte Carlo avec le calcul en nombres flous (méthode des alpha-coupes) L'approche est décrite ci-dessous.

On considère un modèle M faisant intervenir un certain nombre de paramètres dont certains peuvent être représentés par des distributions de probabilité, tandis que d'autres sont jugés mieux représentés par des nombres flous (l'information étant incomplète ou vague). Pour se référer à l'exemple concret traité dans le chapitre 4, on considère un calcul de dose de métal absorbée par une cible humaine.

On a : Dose = $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$

où : P_1 à P_n sont les n variables pouvant être représentées par des distributions de probabilité ; F_1 à F_m sont les m variables pouvant être représentées par des nombres flous.

On supposera que les n variables P sont indépendantes.

Le calcul « hybride » consiste à :

1) générer n nombres aléatoires et échantillonner les n distributions de probabilité P_1 *à* P_n. On obtient un échantillon (P_{x1}, ...P_{xn}),
2) choisir une valeur α de la fonction d'appartenance des nombres flous (en pratique on commence à la valeur 0),
3) sélectionner, pour chaque nombre flou F_1, ... F_m, les valeurs de F_i correspondant à cette valeur de α (α-coupes),
4) chercher les valeurs *min* et *max* de $M(P_{x1}, \ldots, P_{xn}, F_1, \ldots, F_m)$ sur les α-coupe des F_i,
5) affecter ces valeurs *min* et *max* aux limites inférieure et supérieure de la α-coupe de $M(P_{x1}, \ldots, P_{xn}, F_1, \ldots, F_m)$ pour cette valeur de α,
6) répéter l'opération pour une autre valeur de α (on retourne au pas 2). En pratique on procède par pas de 0.1 pour aller de 0 à 1.
7) reconstituer le résultat flou $M(P_{x1}, \ldots, P_{xn}, F_1, \ldots, F_m)$ à l'aide des valeurs min et max de chaque α-coupe de $M(P_{x1}, \ldots, P_{xn}, F_1, \ldots, F_m)$,
8) générer de nouveau n nombres aléatoires et échantillonner les n distributions de probabilité. On obtient un nouvel échantillon (P_{y1}, ...P_{yn}),
9) recommencer le calcul en nombres flous.
10) poursuivre le processus jusqu'à obtenir un nombre « suffisant » de résultats flous.

En résumé, le processus consiste à générer ω doses calculées, chacune étant représentée par un nombre flou. Ceci est représenté de manière schématique par la Figure 16. A noter que l'approche proposée est analogue à un calcul d'intervalle flou (Dubois et Prade, 1991).

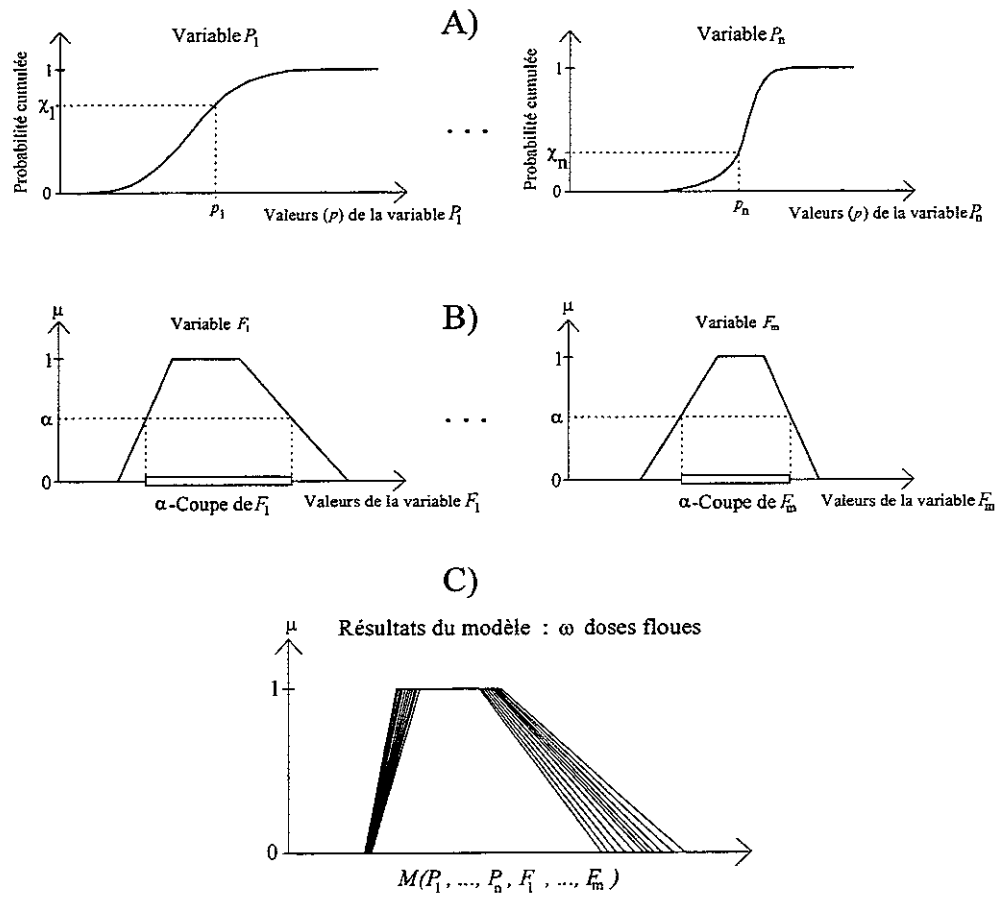


Fig. 16 – Représentation schématique du calcul «hybride »

A) Générer n nombres aléatoires (χ) pour échantillonner les n distributions de probabilité : on obtient n valeurs $p_1, \ldots, p_n$.
B) Appliquer la méthode des α-coupes : chercher les valeurs *Inf et Sup* de $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$ sur les α-coupes. Construire le nombre flou de $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$ reporté en C).
C) Répéter la procédure ω fois : on obtient ω résultats flous de $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$

La question se pose ensuite de la représentation finale de ce calcul sous la forme d'un résultat unique. Une alternative consiste à prendre l'enveloppe de tous les résultats flous. Cette approche, excessivement sécuritaire, serait analogue à supposer la totale dépendance entre les différentes variables aléatoires. Elle ne tient donc pas compte de la multiplicité des probabilités, un scénario combinant des valeurs peu probables des paramètres d'entrée du modèle ayant très peu de chances d'être réalisé.

Une autre alternative, jugée plus réaliste, consiste à dresser des diagrammes de fréquence cumulée pour les valeurs *min* et *max* des réalisations de $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$ pour chaque valeur de α, et à sélectionner les valeur *min* et *max* finales pour un certain degré de confiance (par exemple 95%). Cette approche est cohérente avec le fait que la dispersion de ces valeurs *min* et *max* est entièrement imputable à l'échantillonnage aléatoire probabiliste. Elle est illustrée par la Figure 17.

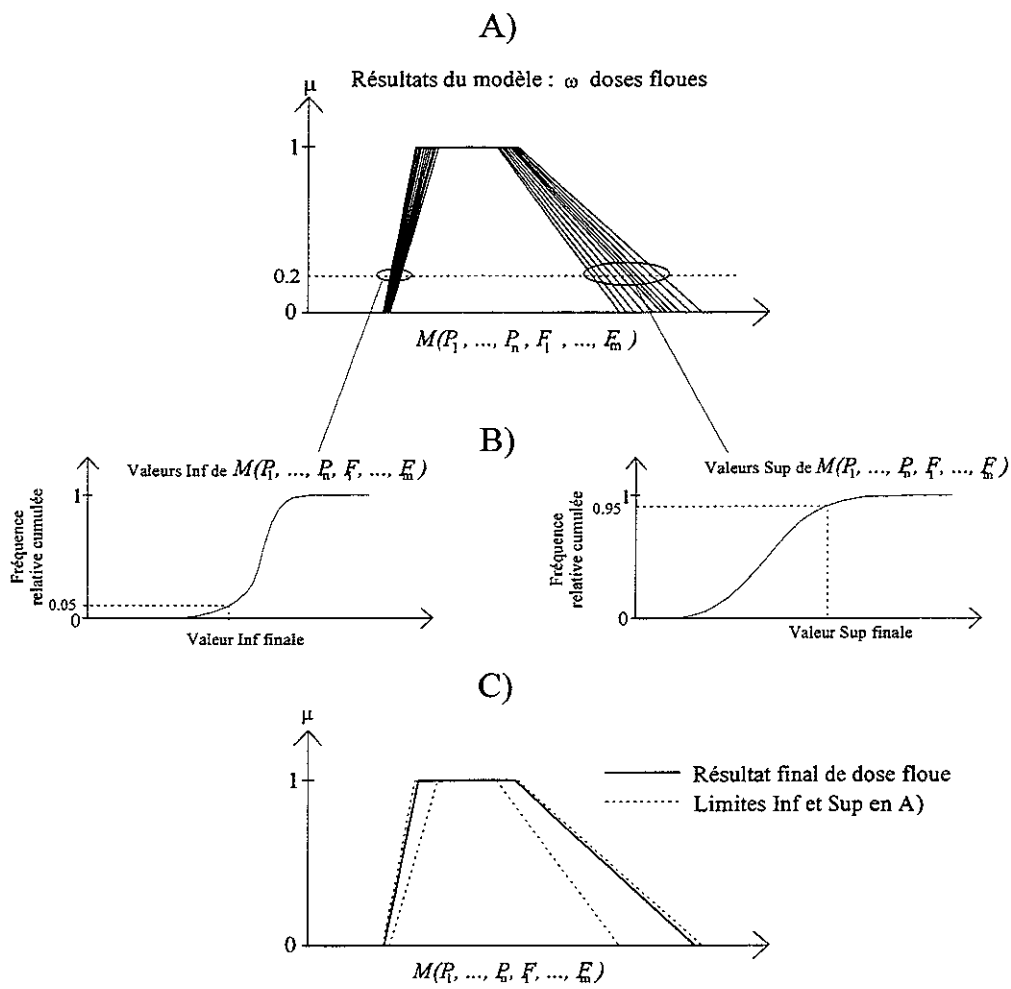


Fig. 17 – Méthode pour extraire le résultat final

A) Sélectionner une valeur α et construire le diagramme des fréquences relatives cumulées des valeurs *Inf* et *Sup* de $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$.

B) Sélectionner les valeurs finales de *Inf* et *Sup* de $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$ pour un degré de confiance de 95%.

C) Répéter la procédure pour chaque valeur de α pour obtenir le résultat flou final de $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$.

On soulignera que cette méthode « hybride » permet de représenter l'incertitude de manière cohérente par rapport à l'information dont on dispose, et éviter notamment de tirer des distributions de probabilité « d'un chapeau » lorsqu'il n'y a pas suffisamment d'information pour les justifier. Par ailleurs, parce qu'elle combine le calcul Monte Carlo avec le calcul par nombres flous, la méthode « hybride » permet de bénéficier de l'information « riche » fournie par les distributions de probabilité, tout en conservant le caractère sécuritaire du calcul flou. Elle présente donc une utilité en terme d'une « application raisonnable » du principe de précaution. Ce dernier point est illustré par le cas concret présenté dans le chapitre suivant.

4. Application au cas d'un sol pollué par du cadmium

4.1. DONNEES MESUREES

Le site étudié est celui de l'usine métallurgique de Métalleurop à Noyelles-Godault. Les sols superficiels de ce site, qui font l'objet d'investigations depuis les années 70, présentent localement des teneurs élevées en certains métaux (Pb, Cd, ..), en raison notamment de retombées liées aux fumées émises par une cheminée implantée sur le site de l'usine. Des teneurs totales en Cd mesurées dans les sols superficiels (0-10 cm) de ce site sont présentées dans le Tableau 5. Les sols ont été analysés par spectrophotométrie d'absorption atomique après mise en solution par de l'acide fluorhydrique. Chaque prélèvement de sol est référencé par coordonnées Lambert. Le nombre important de valeurs permet une interprétation de la variabilité spatiale de la teneur en cadmium du sol (voir sous-chapitre suivant).

Un avantage important de ce site par rapport à l'objectif de la présente étude est l'existence de mesures de Cd dans des plantes (légumes) cultivées directement sur site (Luttringer et Cormis, 1979). Des valeurs de cadmium mesurées dans les parties comestibles de poireaux sont présentées dans le Tableau 6. Les analyses ont été effectuées après un lavage identique à celui pratiqué par une ménagère en vue de la consommation des produits. Le cadmium a été mesuré après attaque des échantillons, préalablement calcinés, par de l'acide chlorhydrique puis par HCO_4-HF. Bien que ces mesures soient très peu nombreuses, elles permettent néanmoins de se rapprocher de la réalité de manière beaucoup plus représentative qu'à partir d'un modèle de calcul. En effet, comme il a été montré dans le chapitre 2, il n'existe pas de modèle fiable permettant de déduire les quantités de métal absorbé par la plante, sur la base des teneurs totales dans les sols.

Tableau 5 – Teneurs totales en Cd dans les sols superficiels du site (0-10 cm)

No. point	ppm Cd	No. point	ppm Cd	No. point	ppm Cd	No. point	ppm Cd
1	16.7	32	2.0	63	2.7	94	0.9
2	8.2	33	1.6	64	0.4	95	0.8
3	5.2	34	1.8	65	1.0	96	8.7
4	3.5	35	1.3	66	1.2	97	5.7
5	0.7	36	10.5	67	11.6	98	3.2
6	1.3	37	4.0	68	5.1	99	1.1
7	0.8	38	3.4	69	2.8	100	0.0
8	0.7	39	3.1	70	1.0	101	6.1
9	7.6	40	7.7	71	0.9	102	3.4
10	4.1	41	2.3	72	0.6	103	4.3
11	3.8	42	1.8	73	18.2	104	1.1
12	5.6	43	27.4	74	6.2	105	0.7
13	0.8	44	19.1	75	4.5	106	0.7
14	7.7	45	4.1	76	2.5	107	8.6
15	5.4	46	3.6	77	1.7	108	3.2
16	3.6	47	2.6	78	1.1	109	2.2
17	0.7	48	3.1	79	0.9	110	0.9
18	0.6	49	7.7	80	0.8	111	0.8
19	1.1	50	8.6	81	13.8	112	0.6
20	20.4	51	1.7	82	8.3	113	5.7
21	14.3	52	7.8	83	4.1	114	3.2
22	5.7	53	4.5	84	3.1	115	3.2
23	4.0	54	4.3	85	1.7	116	3.4
24	2.7	55	2.8	86	0.9	117	0.6
25	2.1	56	5.0	87	0.9	118	9.1
26	0.6	57	1.4	88	0.6	119	4.5
27	1.2	58	2.0	89	15.6	120	3.9
28	12.3	59	16.6	90	11.3	121	2.3
29	7.0	60	7.4	91	8.1	122	0.9
30	3.8	61	4.7	92	4.3	123	0.8
31	3.8	62	3.8	93	2.4	124	1.3

Tableau 6 – Valeurs mesurées dans les poireaux (lavés) (Luttringer et Cormis, 1979)

Sols (mg/kg sec)	2.6	4.3	5.4	8.3	11.6
Poireaux lavés (mg/kg sec)	0.9	1.7	1.9	1.7	2

4.2. ANALYSE GEOSTATISTIQUE

L'analyse de la répartition des 124 teneurs en Cd dans les sols révèle une décroissance au fur et à mesure que l'on s'éloigne de la cheminée d'usine, source de la pollution. On établit donc une régression entre la teneur en Cd et la distance à la cheminée (Figure 18). Plus exactement, cette régression est calculée sur les logarithmes des teneurs. En chaque point de mesure on dispose ainsi du logarithme de la teneur en Cd mesurée, LCd, ainsi que de la valeur de la régression compte tenu de la distance échantillon-cheminée.

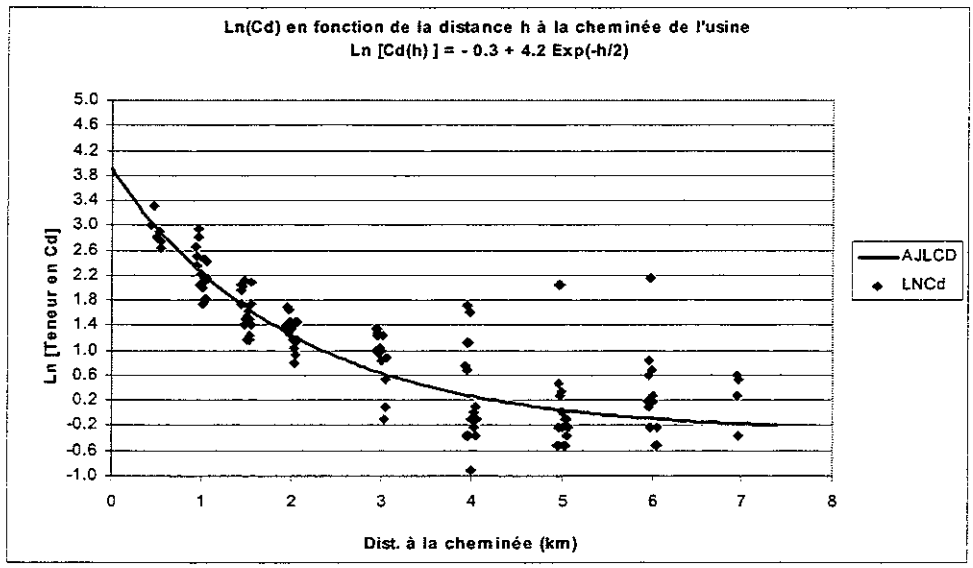


Fig. 18 – Corrélation entre le logarithme de la teneur en Cd et la distance entre le point de mesure et la cheminée de l'usine

Soit LCd* cette valeur. A partir de ces deux valeurs, un résidu R = LCd-LCd* est calculé. On notera que le fait de travailler en logarithme permet d'obtenir un résidu R relativement indépendant de l'échelle des teneurs. Il est ensuite possible d'étudier la répartition spatiale de ce résidu. Le variogramme de R apparaît isotrope (Figure 19) et le variogramme toutes directions confondues peut être ajusté par un modèle linéaire (Figure 20).

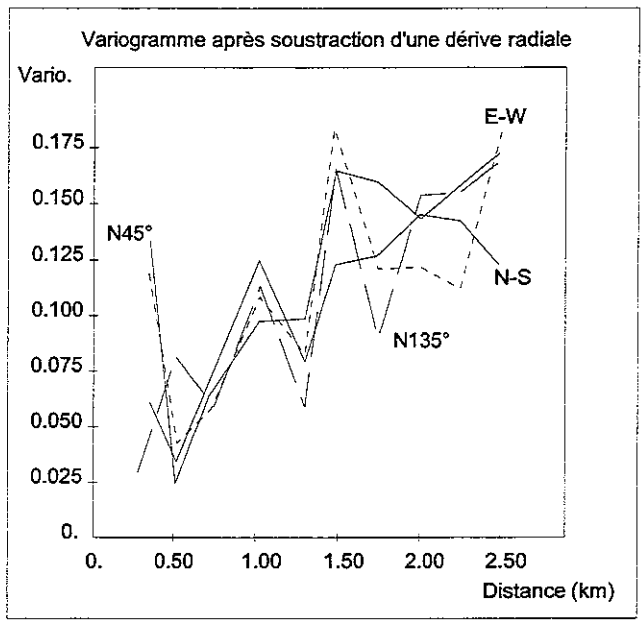


Fig. 19 - Variogramme des résidus après soustraction d'une dérive. 4 directions de calcul (N-S, E-W, N45°, N135°). Ce variogramme est isotrope

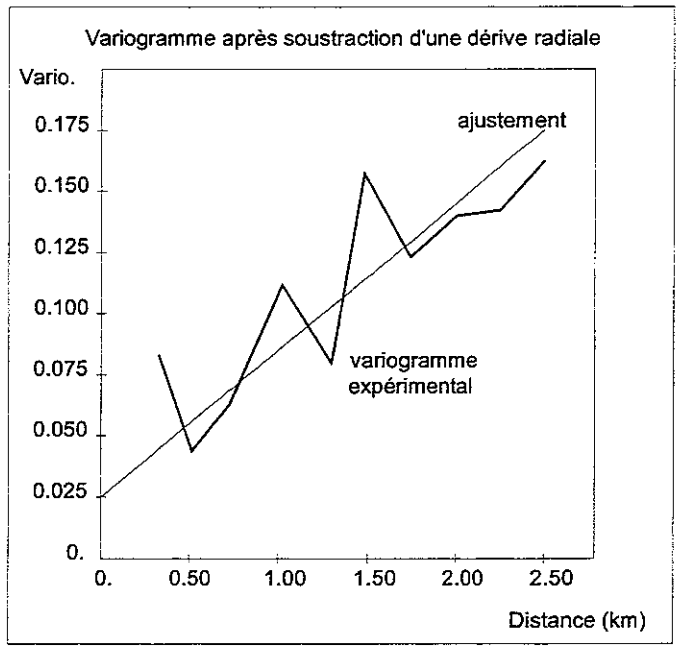


Fig. 20 - Variogramme toutes directions confondues des résidus après soustraction d'une dérive Ajustement par un schéma linéaire : $\gamma(h) = 0.025 + 0.06\ h$

Les teneurs en Cd (leur logarithme) peuvent alors être interpolées par krigeage, méthode qui présente l'avantage de pouvoir calculer également l'écart-type de krigeage, c'est-à-dire un ordre de grandeur de la précision avec laquelle l'estimation est réalisée. En rajoutant la dérive et en prenant l'exponentielle du résultat, on revient aux teneurs

brutes. Ceci permet de tracer la carte des teneurs en Cd (Figure 21), ainsi que la carte des erreurs d'estimation des teneurs (Figure 22). Par rapport à un krigeage direct, sans prise en compte de la dérive, cette méthode permet d'obtenir des erreurs d'estimation plus réalistes.

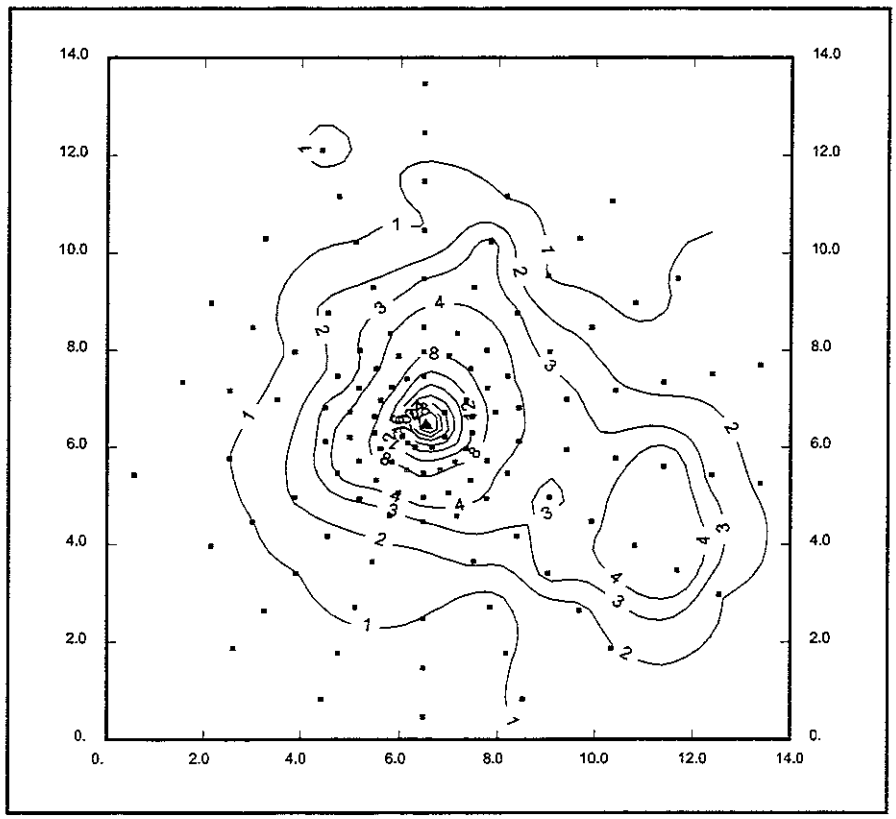


Fig. 21 – Carte de répartition des teneurs en Cd (ppm) des sols superficiels

La carte de répartition des erreurs fait apparaître des « creux », qui correspondent aux points de mesure. En effet, là où la teneur a été directement mesurée, l'erreur sur l'estimation est minimale (et correspond à l'erreur de mesure). En chaque point du maillage, la teneur en Cd du sol peut donc être représentée par une valeur moyenne (la valeur estimée ; Figure 21) et par un écart type (obtenu à partir de la carte des erreurs sur l'estimation ; Figure 22). On a donc une représentation « statistique » de la teneur en Cd en chaque point du maillage.

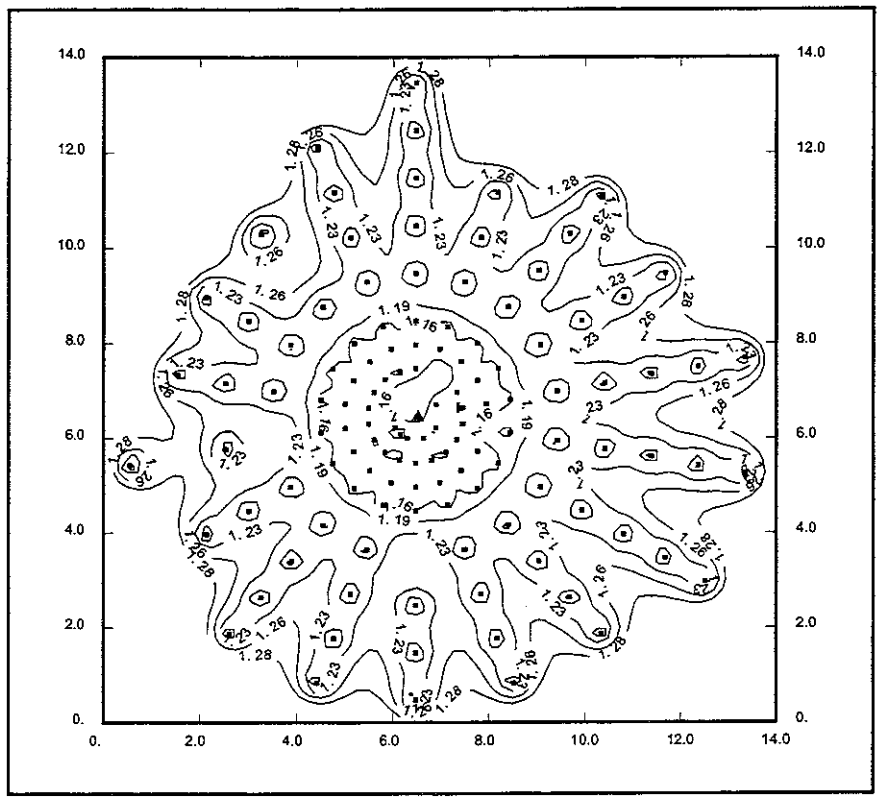


Fig. 22 – Carte de répartion de l'erreur sur l'estimation de la teneur en Cd

4.3. ANALYSE CORRELATOIRE FLOUE

Les mesures de teneurs dans les plantes (Tableau 6) ne sont pas en nombre suffisant pour permettre une analyse statistique. Une analyse de type possibiliste se prête bien plus à l'information dont on dispose dans ce cas.

Dans un premier temps, on a tenté de délimiter les données mesurées par des droites de régression en établissant des bornes internes, et une borne supérieure jugée sécuritaire (Figure 23).

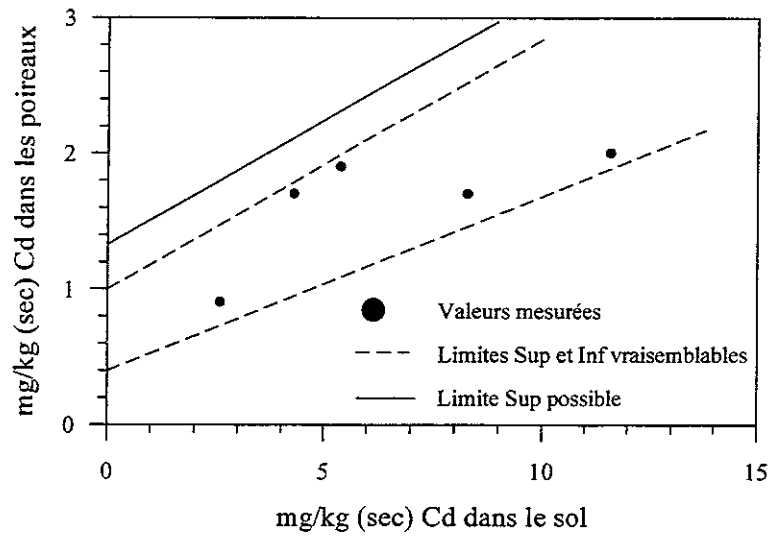


Fig. 23 – Encadrement des données mesurées dans les poireaux par des droites

Mais un défaut important de cette représentation est qu'elle ne tient pas compte de l'existence d'un seuil d'absorption par la plante : elle suppose qu'au fur et à mesure qu'augmente la teneur dans le sol, la teneur dans la plante augmente proportionnellement. Une représentation jugée plus réaliste est celle de la Figure 24.

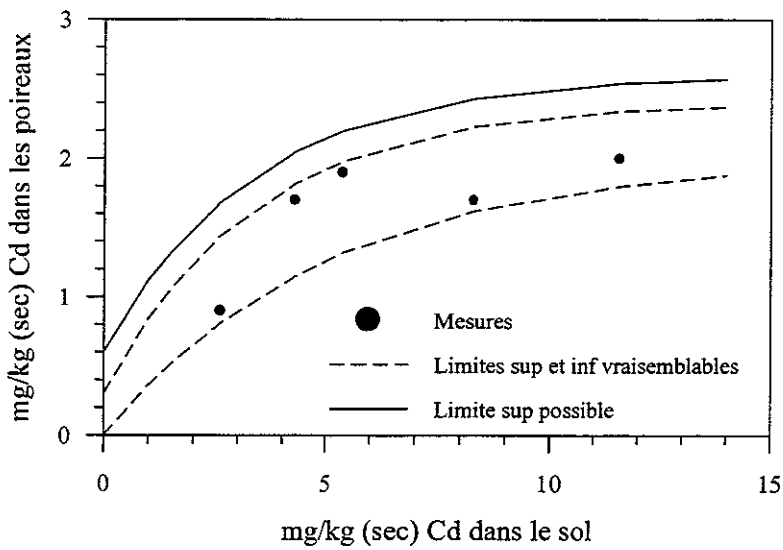


Fig. 24 – Encadrement des données mesurées dans les poireaux par des courbes.

Les courbes en pointillés de la Figure 24 signifient que si on mesure la concentration dans des poireaux cultivés sur ce site, il est vraisemblable qu'on obtienne une valeur située entre ces deux courbes. La courbe en trait plein représente une limite supérieure jugée possible pour cette concentration. Tandis que l'on peut débattre sur la position précise de cette courbe haute, on notera que ce mode de représentation est cohérent avec

les quelques données disponibles, et qu'elle peut être facilement modifiée pour prendre en compte l'avis d'experts en sciences agronomiques.

A partir de la Figure 24, on établit une corrélation entre la teneur en cadmium dans les poireaux, et celle dans le sol :

$$Cd_{pl}^{*} = Or + (As - Or)\left(1 - \exp\left(-k\, Cd_{s}\right)\right) \qquad (15)$$

où : Cd_{pl}^{*} = concentration en cadmium dans la partie comestible de la plante (mg Cd / kg plante sèche) ; Cd_s = concentration en cadmium mesurée dans le sol (mg Cd / kg sol sec) ; *Or* = valeur de Cd_{pl}^{*} à l'origine (Cd_s = 0) ; *As* = Asymptote (Cd_{pl}^{*} pour des valeurs élevées de Cd_s) ; *k* = paramètre contrôlant la forme de la courbure dans la Figure 24.

Les paramètres *Or*, *As* et *k* dans l'équation (15) sont représentés par des nombres flous, tandis que les valeurs de Cd_s sont représentées par des fonctions de distribution de probabilité qui découlent de l'analyse géostatistique (voir sous-chapitre précédent). En s'appuyant de nouveau sur la Figure 24, on identifie les nombres flous qui sont représentés dans la Figure 25 (ainsi qu'un autre paramètre décrit plus bas).

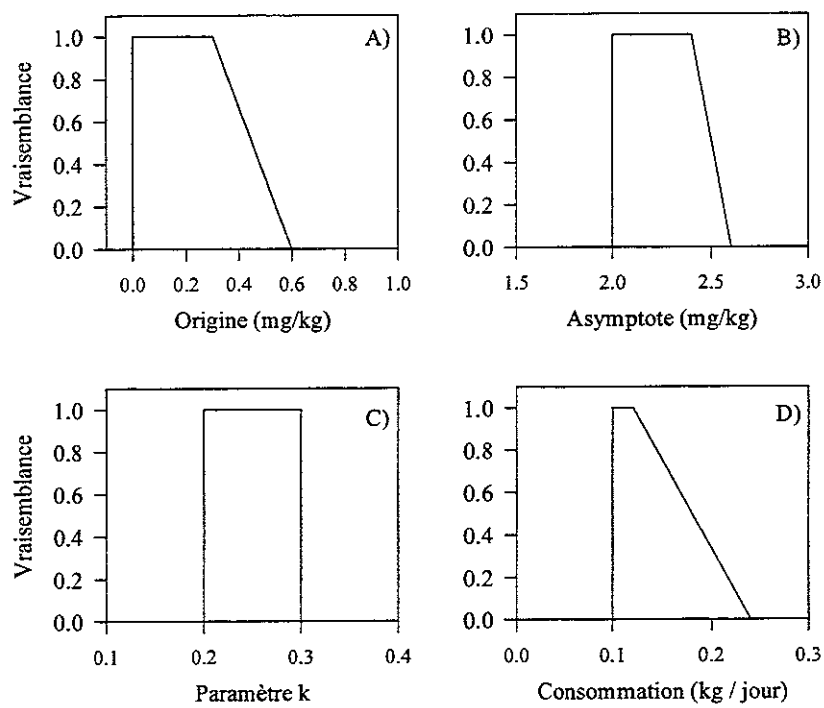


Fig. 25 – Nombres flous pour plusieurs paramètres

Par exemple le nombre flou en Figure 25.B) signifie que la valeur maximale de cadmium est vraisemblablement située entre 2 et 2.4 mg/kg, tandis qu'une valeur maximale de 2.6 mg/kg est considérée comme étant possible.

4.4. CALCUL DU RISQUE D'EXPOSITION ET REPRESENTATION GRAPHIQUE

Le calcul de dose est effectué à partir de l'équation suivante :

$$Dose = \frac{Cd_{pl}^{*} \cdot 1000 \cdot Con \cdot DMC}{BW} \qquad (16)$$

où : Dose = dose absorbée par la cible humaine (µg Cd par jour et par kg de poids corporel) ; *Con* = consommation journalière de poireaux (kg par jour) ; *DMC* = teneur en matière sèche des poireaux (pourcentage pondéral) ; *BW* = Poids corporel de la cible humaine (kg).

La teneur en matière sèche des poireaux tient compte du fait que la consommation journalière est définie par rapport à la masse totale humide, tandis que la concentration en cadmium dans les poireaux est définie par rapport à la masse sèche. La concentration en cadmium dans les poireaux (Cd_{pl}^{*}) obtenue à l'aide de l'équation de corrélation floue (équation 15).

Il est supposé (hypothèse simplificatrice) que les poireaux sont représentatifs des légumes pour ce qui concerne l'exposition de l'homme au cadmium par le biais de la consommation de légumes. La consommation journalière de légumes et la teneur en matière sèche sont sélectionnées sur la base de données proposées par INERIS (1999). La consommation vraisemblable est prise entre 100 et 120 g/jour, tandis qu'une valeur haute possible est prise comme étant égale à deux fois la valeur haute. (240 g/jour). La teneur en matière sèche (*DMC*) et le poids corporel sont considérés comme étant constants : 15% et 70 kg respectivement.

Un exemple de résultat du calcul hybride pour Cd_s moyen = 7.97 ppm, avec une erreur de 1.15, est présenté en Figure 26.

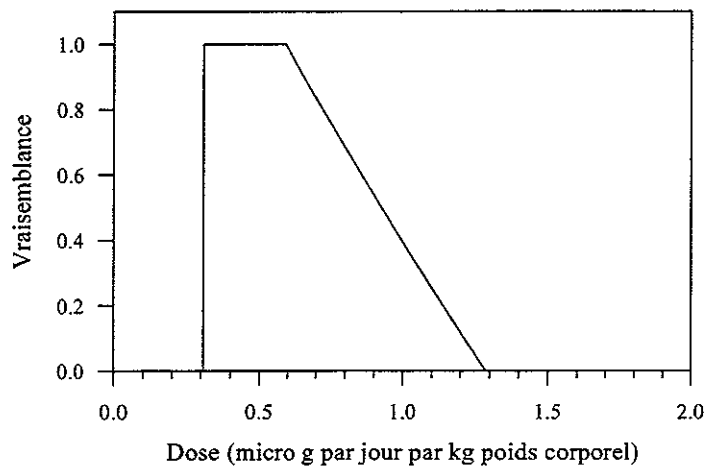


Fig. 26 – Résultat de calcul de dose pour Cd_s moyen = 7.97 ppm et erreur = 1.15

On s'intéresse ici à la **possibilité de dépassement d'une dose jugée tolérable**. On choisit comme dose journalière tolérable la valeur de 1 $\mu g/kg\,j^{-1}$ préconisée par l'OMS (OMS, 1994).

En appliquant la procédure décrite dans le sous-chapitre 3.4 on obtient une carte de possibilité de dépassement de la dose tolérable qui est présentée dans la Figure 27.

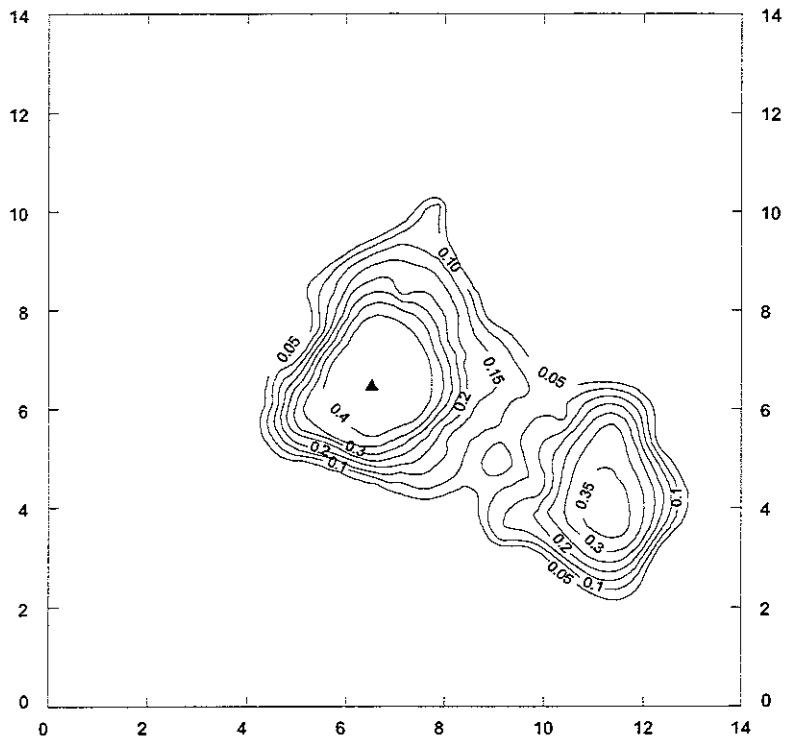


Fig. 27 – Carte de la « possibilité » que la dose absorbée suite à la consommation de légumes dépasse la dose jugée tolérable (1 $\mu g/kg\,j^{-1}$)

On distingue des zones où il y a une forte possibilité de dépassement (les zones proches de la cheminée d'usine au centre) et des zones où cette possibilité est faible voire nulle.

Ce type de représentation constitue un outil d'aide à la décision en matière de gestion de la contamination du site. A noter **qu'il appartiendra à l'autorité sanitaire compétente de fixer une limite pour le niveau de possibilité de dépassement de la dose tolérable**. Des valeurs de 0.1 ou 0.2, par exemple, se traduisent par une restriction sur la culture de légumes destinés à la consommation humaine, sur une superficie d'environ 20 kilomètres carrés. A noter qu'exiger une possibilité de zéro peut, dans de nombreuses situations, être trop strict et se traduire par des restrictions d'usage sur des superficies de terrain trop importantes, ou par des coûts de réhabilitation trop élevés. Comme le calcul en nombres flous est plus majorant que le calcul Monte Carlo (voir sous-chapitre suivant), les critères de tolérance en terme de possibilités ne doivent pas être aussi stricts qu'en terme de probabilités résultant d'une analyse Bayésienne.

4.5. CONSEQUENCES DE L'HYPOTHESE DE DISTRIBUTIONS DE PROBABILITE A PRIORI

Pour illustrer les possibles conséquences de l'hypothèse de distributions de probabilité, les nombres flous de la Figure 25 ont été convertis en fonctions de densité de probabilité, puis en fonctions de distribution, en ajustant les axes des Y pour que les intégrales sous les courbes soient égales à 1. On notera que cette procédure ne constitue pas la manière rigoureuse d'établir une correspondance entre des distributions de probabilité et des nombres flous, mais qu'elle « simule » l'hypothèse d'assigner des probabilités aux valeurs limites plutôt que des possibilités.

Pour le calcul, une valeur élevée de cadmium dans le sol a été sélectionnée (Cd_s = 25 ppm, erreur = 1.2) c'est à dire une valeur située tout près de la cheminée. Les distributions de probabilité et les résultats de la simulation Monte Carlo (1000 itérations) sont présentées dans la Figure 28. Dans la Figure 29, la fréquence relative cumulée est comparée à la dose de référence.

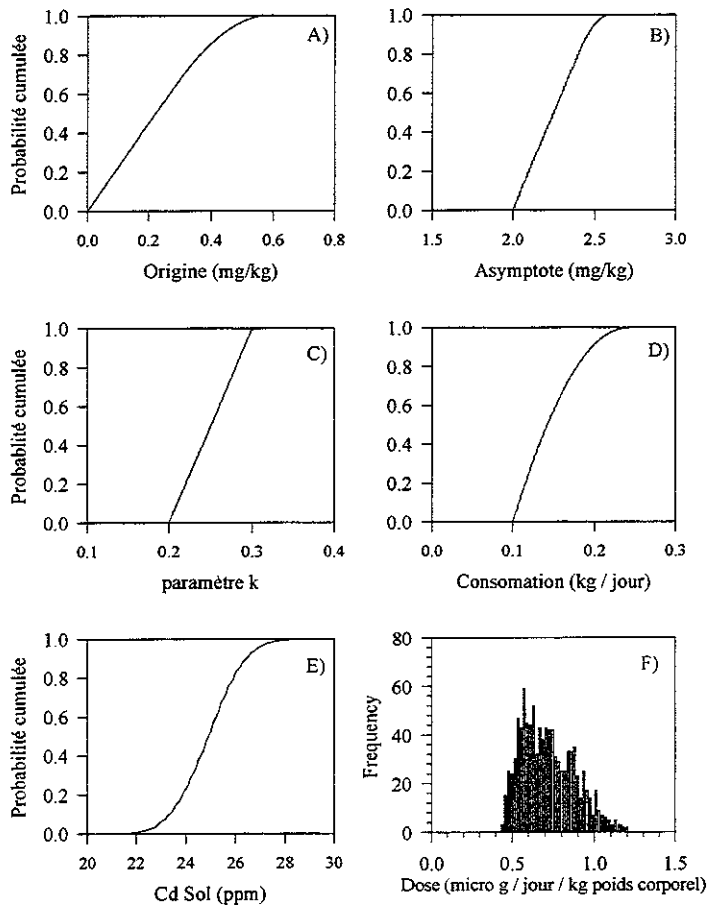


Fig. 28 – PDF utilisés pour la simulation Monte Carlo et résultat (F)

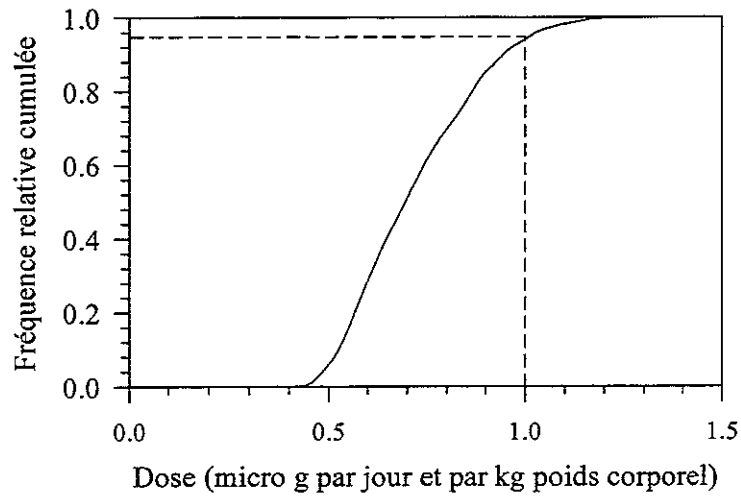


Fig. 29 – Comparaison du résultat de la simulation Monte Carlo avec la dose jugée tolérable (1 $\mu g/kg\,j^{-1}$)

On constate dans la Figure 29 que la dose calculée est jugée tolérable, puisqu'il y a une probabilité inférieure à 5% d'avoir une valeur supérieure à la dose maximale admissible (1 µg/kg j^{-1}). Comme la valeur de Cd_{sol} utilisée correspond à une valeur mesurée très près de la cheminée, on en déduirait que tout le site serait « apte » à la culture de légumes à des fins de consommation. Cet exemple illustre donc bien le caractère non-sécuritaire de l'hypothèse des probabilités subjectives.

5. Traitement du problème « inverse »

5.1. EXPRESSION DU PROBLEME INVERSE

On s'intéresse ici à la question suivante : **comment définir la concentration du sol C_s de manière à ce que la dose estimée ne dépasse pas la dose jugée tolérable (notée D_0) ?**

Compte tenu de l'incertitude liée aux différents paramètres du modèle de calcul de la dose, cette question peut s'exprimer dans un cadre possibiliste de la manière suivante :

« Comment définir la concentration du sol Cd_s de manière à ce que la possibilité (notée Π) que la dose calculée dépasse la dose tolérable D_0 ne soit pas supérieure à une certaine valeur ? ».

Cette formulation du problème est illustrée graphiquement dans la Figure 2 ci-dessous (voir Prade et Dubois, 1988). La dose calculée est un nombre flou représenté par sa fonction de répartition (μ). La possibilité de dépassement de la dose tolérable (D_0, considérée ici comme étant un nombre précis) est ici une valeur notée α.

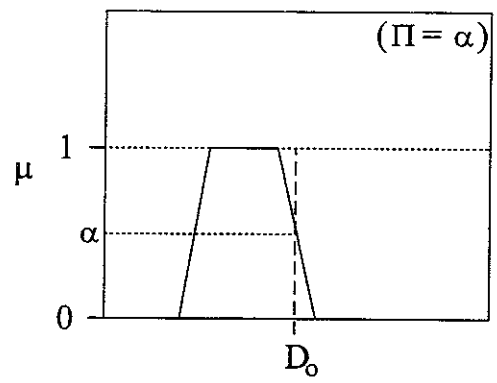


Fig. 30 – Comparaison possibiliste d'une dose calculée et d'un seuil D_0.
La possibilité de dépassement du seuil est ici : $\Pi = \alpha$.

5.2. METHODOLOGIE GENERALE

Ce problème inverse peut être traité de manière générale à l'aide d'une approche itérative. L'approche hybride définie dans le chapitre 3 peut être mise en œuvre en faisant varier Cd_s jusqu'à ce que la possibilité de dépassement de la dose tolérable D_0 par la dose floue calculée corresponde à la valeur souhaitée. Mais cette approche peut

s'avérer être très lourde en temps de calcul, aussi dans les cas simples est-il préférable d'adopter une approche spécifique

5.3. METHODOLOGIE SPECIFIQUE

Cette méthodologie est illustrée à partir du même calcul de dose que dans le chapitre 4. L'examen des Equations (15) et (16) montre que la dose calculée atteint son maximum lorsque les paramètres *Or*, *As*, *k* et *Con* sont maximals. Ces valeurs maximales, en fonction du degré de vraisemblance α, sont :

$Or_{max} = 0.6 - 0.3\ \alpha$
$As_{max} = 2.6 - 0.2\ \alpha$
$k_{max} = 0.3$
$Con_{max} = 0.24 - 0.12\ \alpha$

A noter que pour un nombre flou de forme trapézoïdale, de support [a, b] et de noyau [c, d], la valeur maximale en fonction du degré de vraisemblance α est : d – (d – b) α.

La dose maximale est donc (17) :

$$Dose_{max} = DMC \cdot 1000 \cdot (0.24 - 0.12\alpha) \cdot [0.6 - 0.3\alpha + (2 + 0.1\alpha)(1 - \exp[-0.3C_s])] / BW$$

A partir de l'équation (17), on peut exprimer Cd_s en fonction de α en fixant : $Dose_{max} = D_o$, la limite jugée tolérable. On obtient :

$$Cd_s = \ln\left[1 + \frac{(0.6 - 0.3\alpha)}{(2 + 0.1\alpha)} - \frac{D_o}{(0.24 - 0.12\alpha)\lambda(2 + 0.1\alpha)}\right] / (-0.3)$$

où : $\lambda = MS^*1000/BW$

Pour une dose journalière maximale recommandée de 1 µg / j kg^{-1} (OMS, 1994), on obtient le graphique de la Figure 31.

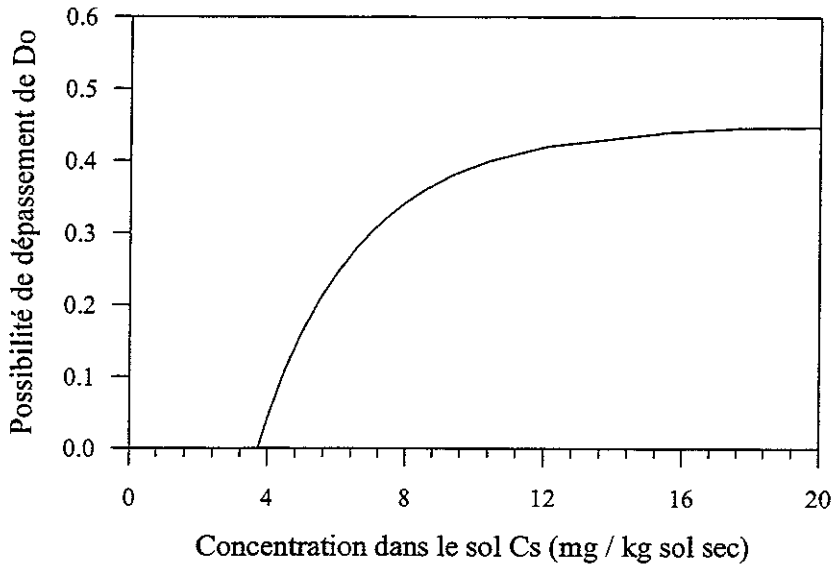


Fig. 31 – Concentrations en Cd dans le sol correspondant à différentes possibilités de dépassement de la dose jugée tolérable (1 µg / j kg^{-1}).

On note en Figure 31 qu'en dessous d'une concentration dans le sol d'environ 4 mg/kg, un dépassement de la dose tolérable est jugée impossible. Au fur et à mesure que la concentration dans le sol augmente, la possibilité de dépassement de la dose tolérable augmente également. Mais cette possibilité n'atteint jamais 1 en raison de l'effet de seuil pris en compte sur la base des données mesurées (Figure 24).

5.4. DISCUSSION

La Figure 31 présente la possibilité de dépassement d'une dose jugée tolérable. On notera **qu'il appartient à l'Autorité Sanitaire compétente de définir la valeur réglementairement acceptable pour le niveau de possibilité d'excéder cette dose tolérable**.

Il est rappelé que dans un cadre possibiliste, exiger une possibilité nulle d'excéder la dose tolérable aboutirait à des objectifs de dépollution extrêmement stricts, peut-être difficilement atteignables à un coût raisonnable. Vollmer et al. (1995) ont proposé une valeur de 0.3 pour établir un critère de concentration en Cd dans des sols superficiels. Comme l'approche adoptée par ces auteurs est très semblable à celle utilisée ici, elle est résumée ci-après.

Sur la base de considérations écotoxicologiques, Vollmer et al. (1995) établissent les propositions suivantes :

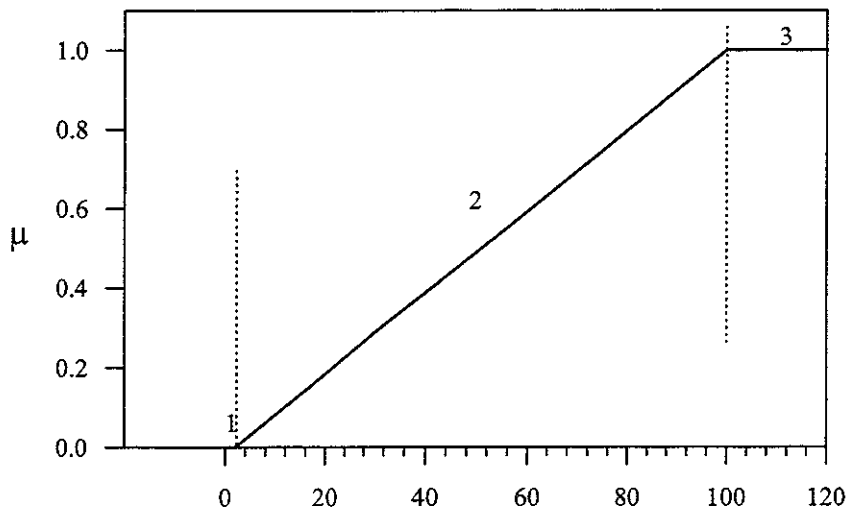
- Aucune nuisance aux plantes et animaux de ferme, ni à l'être humain, n'est constatée pour des concentrations en cadmium inférieures ou égales à 2 mg/kg de sol sec.

- Il y a quasi-certitude d'effets sur la santé des animaux d'élevage pour des concentrations supérieures ou égales à 150 mg/kg.

- Il y a quasi-certitude d'effets sur la santé humaine pour des concentrations de l'ordre de 100 mg/kg.

Une valeur de qualité du sol en cadmium, dite "d'intervention", protégeant le récepteur le plus sensible (l'homme) est donc située entre 2 et 100 mg/kg.

Vollmer et al. font l'hypothèse que la possibilité d'effets nuisibles croît linéairement entre la valeur de 2 mg/kg (possibilité nulle) et celle de 100 mg/kg (possibilité totale, fixée à 1). Le critère de qualité du sol est alors représenté, en nombre flou, par la Figure 32.



1 : Zone d'acceptabilité totale
3 : Zone d'inacceptabilité totale
2 : Zone intermédiaire

Fig. 32 – Critère de qualité flou d'un sol (Vollmer et al., 1995).

L'ordonnée de cette figure permet de déterminer une valeur jugée pratiquement acceptable. Vollmer et al. (1995) indique que le niveau de possibilité d'effets nuisibles peut être fixé à 0,3. Ceci correspond à un seuil de teneur en Cd dans le sol de 30 mg/kg.

A noter qu'il s'agit d'un objectif générique, ne tenant pas compte des spécificités locales éventuelles. Cette approche est elle-même dérivée d'un travail antérieur concernant la détermination, en nombres flous, de valeurs limites d'émissions de polluants atmosphériques dont le SO_2 (Pohl et al., 1995).

La démarche suivie comporte plusieurs étapes :

- Construire la fonction d'appartenance de la concentration en SO_2 dans la perspective de la santé humaine : pour cette dernière, il est admis qu'une concentration en SO_2 dans l'air inférieure ou au plus égale à 5 µg/m3 rend impossible les nuisances à la santé, et une valeur supérieure à 200 µg/m3 correspond à une quasi-certitude d'effets nuisibles.
- Prendre en compte de la même façon les aspects liés aux nuisances sur les plantes (ces dernières étant plus sensibles au SO_2 que les humains).
- A l'aide d'un "opérateur de pondération", combiner ces deux fonctions d'appartenance en une courbe unique, assurant le meilleur compromis entre ces deux aspects ; cet opérateur est choisi de manière à ce que la concentration limite actuelle réglementaire (en Suisse), soit 30 µg/m^3, soit obtenue pour une fonction d'appartenance (= possibilité d'effets indésirables) de 0,5 (il est reconnu que cette dernière valeur est **arbitraire**).
- Utiliser la fonction d'appartenance ainsi déterminée pour la concentration en SO_2 dans l'air pour calculer un "volume (d'air) critique flou", c'est-à-dire le volume d'air dans lequel il faut dissoudre 1 g de SO_2 pour garantir la concentration floue tolérable.

Dans ce travail, on utilise donc le fait que la fonction d'appartenance de la concentration dans l'air représente la mesure de possibilité qu'une valeur de concentration induise un effet sanitaire adverse.

6. Conclusions et perspectives

Le projet sur la prise en compte de l'incertitude dans l'évaluation du risque d'exposition de l'homme au polluants du sol, réalisé dans le cadre du Programme GESSOL piloté par l'INRA pour le compte du Ministère chargé de l'Environnement, a permis de mettre au point une méthodologie qui présente des avantages pratiques importants. Elle permet, en effet, de prendre en compte l'incertitude liée aux valeurs des paramètres qui influencent l'exposition de l'homme aux polluants du sols, de manière cohérente par rapport à l'information dont on dispose dans la pratique. Elle se démarque de l'approche dite « Bayésienne », plus communément utilisée, en ce qu'elle s'attache à ne pas introduire artificiellement plus de connaissance qu'il n'en existe réellement. Elle fournit un pont entre les approches probabiliste et possibiliste et permet de bénéficier de leurs avantages respectifs.

Malgré ces atouts, on citera deux inconvénients de la méthode proposée. Le premier est d'ordre opérationnel. La méthode « hybride » combine deux techniques de calcul qui sont gourmandes en temps de calcul : la méthode dite Monte Carlo, et le calcul en nombres flous. Le modèle utilisé dans l'exemple traité dans le chapitre 4 était simple (calcul d'une dose liée à la consommation de légumes). Dans le cas de modèles plus complexes, analytiques ou numériques, il est à craindre que l'application de la méthode ne soit limitée par la puissance des machines disponibles.

L'autre inconvénient que l'on peut citer est d'ordre sociologique. Les décideurs en matière de sols pollués et de risques en général commencent seulement à prendre en compte la notion de « probabilité d'apparition de dommages ». Celle de « possibilité d'apparition de dommages » pourrait rencontrer une certaine réticence de leur part, ou de la part du public et des médias. Il suffit en effet de constater « l'aversion » des médias à l'égard de l'incertitude (« les experts sont payés pour savoir ») pour se rendre compte que la perception sociologique du risque est un aspect auquel il faudra s'intéresser dans l'avenir si l'on souhaite que la démarche de l'évaluation des risques soit utilisée à bon escient.

La méthode « hybride » proposée dans ce rapport pour combiner différents modes de représentation de l'incertitude dans une même évaluation des risques est une alternative parmi d'autres. Comme il est montré en Annexe 1 de ce rapport, il en existe d'autres. Des travaux sont actuellement en cours de démarrage qui viseront à :

- examiner ces autres alternatives,
- déterminer leurs validités et modalités d'application respectives,
- les appliquer à des cas synthétiques et réels.

Références

Baes, C. (1982) – Prediction of radionuclide Kd values from soil-plant concentration ratios. *Trans. Amer. Nucl. Soc.* 41:53-54.

Bonano, E., Hora, S., Keeney, R., and von Winterfeldt, D. (1990). "Elicitation and use of expert judgement in performance assessments for high-level radioactive waste repositories." *Sandia National Laboratories* Report SAND89-1821, Albuquerque, New Mexico.

Bouchon Meunier, B. (1993). La logique floue. *Editions "que sais-je",* Presses Universitaires de France.

Brown, S., Chaney, R., Angle, J., Baker, A. (1995) – Zinc and cadmium uptake of *Thlaspi caerulescens* grown in nutrient solution. *Soil Sci Soc Am J,* 59 :125-133.

Chaney, R., Brow, S., Stuczynski, T., Daniels, W., Henry, C., Li, Y.-M., Siebielec, G, Malik, M., Angle, J., Ryan, J., Compton, H. (1999) – In-situ remediation and phytoextraction of metals from hazardous contaminated soils. *Dans* : Innovative clean-up approaches : Investments in technology developments, results & outlooks for the future. Nov.2-4, Bloomingdale, USA.

Chaney, R. (1985) – Potential effects of sludge-borne heavy metals and toxic organics on soils, plants, and animals, and related regulatory guidelines. Annex 3, Workshop Paper 9, pp.1-56. Dans : Final Report of the Workshop on the International Transportation, Utilization or Disposal of Sewage Sludge Including Recommendations. PNSP/85-01. Pan American Health Organization, Washington, DC.

Chang, A., Page, A., Warneke, J. (1987) – Long-term sludge appliation on cadmium and zinc accumulation in Swiss chard and radish. *Journal of Environmental Quality*, 16, 217-221.

Clarkson, D., Luttge, U. (1989 – Mineral nutrision : divalent cations, transport and compartmentalization. *Prog. Bottany*, 51 : 93-112.

Côme, B., Guyonnet, D., Magnouni, S. (1997) - Apports potentiels de la théorie des possibilités à l'évaluation des risques environnementaux. *Déchets-Sciences et Techniques no.7, pp.40-46.*

Cullen, A.C., and Frey, H.C. (1999). - Probabilistic techniques in exposure assessment: a handbook for dealing with variablility and uncertainty in models and inputs. New York Plenum Press, 333 pp.

Del Castilho, P., Chardon, W. (1995) – Uptake of soil cadmium by three field crops and its prediction by a pH-dependent Freundlich sorption model. *Plant and Soil*, 171 :263-266.

Delgado, M., Moral, S. (1987) - On the concept of possibility-probability consistency . *Fuzzy Sets and Systems* 21, 311-318.

Dijkshoorn, W., Lampe, J., VanBroekhovern L. (1989) – The effect of soil pH and chemical form of nitrogen fertilizer on heavy metal contents in ryegrass. *Fertilizer Research*, 4 :63-74.

Dubois, D., Prade, H. (1982) - On Several Representations of an Uncertain Body of Evidence. *Fuzzy Information and Decision Processes* (M. M. Gupta, E. Sanchez editeurs),167-181.

Dubois, D., Prade, H. (1994) - Unfair Coins and Necessity Measures », *Fuzzy Sets and Systems* 10, 15-20.

Guyonnet, D., Côme, B., Perrochet, P., Parriaux, A. (1999) – Comparing two methods for addressing uncertainty in risk assessments. *Journal of Environmental Engineering*, 125 :7, 660-666.

Heller, R., Esnault, R., Lance, C. (1998) – Physiologie végétale. 1. Nutrition. *Editions Dunod*, Paris.

Hooda, P.S., Alloway, B.J. (1993) – Effects of time and temperature on the bioavailability of Cd and Pb from sludge-amended soils. *Journal of Soil Science*, 44:97-110.

INERIS (1999) – Méthode de calcul des valeurs de constat d'impact dans les sols. INERIS, Verneuil-en-Halatte, France.

Jarvis, S., Jones, L., Hopper, M. (1976) – Cadmium uptake from solution by plants and its transport from roots to shoots. *Plant and Soil*, 44 :179-191.

Jopony, M. (1991) – The solubility of lead and cadmium in contaminated soils. Thèse de l'Université de Nottingham (Royaume Uni).

Kahru, A., Pollumaa, L., Ivask, A., Virta, M., Douay, F., Doubourguier, H.C. (2002) – Biodisponibilité et écotoxicité des métaux lourds dans les sols agricoles pollués par les activités industrielles. *FORUM Qualité de sols*. Ministère de l'Agriculture, 15-16 mai 2002, Paris.

Klir, G. J. (1990) - A principle of uncertainty and information invariance. *International. Journal of General Systems*, 17(2-3), 249-275.

Kumar, P., Dushenkov, V., Motto, H., Raskin, I. (1995) - Phytoextraction – the use of plants to remove heavy metals from soils. *Environ. Sci. Technol.* 29 : 1232-1238.

Lorenz, S., Hamon, R., Holm, P., Domingues, H., Sequeira, E., Christensen, T., McGrath, S. (1997) – Cadmium and zinc in plants and soil solutions from contaminated soils. *Plant and Soil*, 189:21-31.

Luttringer, M., de Cormis, L. (1979) – La pollution par les métaux lourds à Noyelles-Godault et ses environs (Pas de Calais). Rapport INRA.

Mench, M., Fargues, S. (1994) – Metal uptake by iron-efficient and inefficient oats. *Plant and Soil*, 165 :227-233.

Nouzille, C., Nicoullaud, B., Duval, O., Golaz, F., Couturier, A., Bruand, A. (1998) - Etude de l'alimentation naturelle de la nappe de Beauce. Elaboration d'un modèle de calcul de l'infiltration efficace. Rapport INRA, Orléans.

Palm (1994) – A model for sorption, flux and plant uptake of cadmium in a soil profile : model structure and sensitivity analysis. *Water, Air and Soil Pollution* 77 :169-190.

Poels, C.L.M., Gruntz, U., Isnard, P., Riley, D., Spiteller, M., ten Berge, W., Veerkamp, W., Bontinck, W., (1990) – ECETOC technical report nr.40. »Hazard Assessment of Chemical Contaminants in Soil.

OMS (1994) – Directives de qualité pour l'eau de boisson. Volume 1 : Recommandations. 2ème édition. Organisation Mondiale de la Santé, Genève Suisse.

Prade, H., Dubois, D. (1988) - Théorie des possibilités - Application à la représentation des connaissances en informatique. *Collection "Méthode et Programmes", Masson Ed., Paris.*

Pohl, C., Ros M., Waldeck, B., Dinkel, F. (1995) - Fuzzy - Immissionsgrenzwerte. Eine Methodik zur unscharfen Modellierung von Immissionsgrenzwerten. Carbotech AG, Basel (CH).

Salt, D., Blayblock, M., Kumar, N., Dushenkov, V., Ensley, B., Raskin, I. (1995) - Phytoremediation : a novel strategy for the removal of toxic metals from the environment using plants. *Biotechnology*, Vol.13, pp.468-474.

Schnoor, J. (1997) – Phytoremediation. Technology Evaluation Report. TE-98-01. *GWRTAC Series*, Pittsburg, PA.

Shafer, G. (1976) - *A Mathematical Theory of Evidence.* Princeton University Press.

Singh, B., Narwal, R., Jeng, A., Almas, A. (1995) – Crop uptake and extractability of cadmium in soils naturally high in metals at different pH levels. *Commun. Soil Sci. Plant Anal.*, 26(13&14), 2123-2142.

Smets, P. et R. Kennes (1994) - « The transferable belief model », *Artificial Intelligence* 66 : 191-234.

Smilde, K., Van Luit, B., Van Driel, W. (1992) – The extraction by soil and absorption by plants of applied zinc and cadmium. *Plant and Soil*, 143 :233-238.

Smith, S. (1993) – Effect of soil pH on availability to crops of metals in sewage sludge-treated soils. I. Nickel, Copper and Zinc uptake and toxicity to ryegrass. *Environmental Pollution* 85 :321-327.

Tessier, A.P., Campbell, G.C., Bisson, M. (1979) – Sequential extraction procedure for speciation of particulate trace metals. *Analytical Chemistry* 51, pp.844-850.

Thornthwaite, C. & Mather, J. (1955) The water balance. *Climatology Vol. 8 No. 1.* Laboratory of Climatology, Centerton, New Jersey, USA.

Veerkamp, W., Berge, W. (1992) – Hazard assessment of chemical contaminants in soil. ECETOC Technical Report 40 Revised. Appendix III. ECETOC, Bruxelles.

Vollmer M.K., Gupta S.K., Krebs R. (1995) - New standards on contaminated soils in Switzerland. Comparison with Dutch and German quality criteria. Proceedings of "Contaminated Soils", Third International Conference on the Biogeochemistry of Trace Elements, Paris, May 1995.

Vose D. (1996) - Quantitative risk analysis - A guide to Monte-Carlo simulation modelling. *John Wiley & Sons Ltd.*

Wheeler, D., Power, I. (1995) – Comparison of plant uptake and plant toxicity of various ions in wheat. *Plant and Soil*, 172 :167-173.

Wild, A. (1988) – Russell's Soil Conditions and Plant Growth. 11ème édition, Longman, Londres.

Xiong, L., Lu, R. (1993) – Effect of liming on plant accumulation of cadmium under upland or flooded conditions. *Environmental Pollution* 79:199-203.

Zadeh L. (1965) - Fuzzy Sets - *Information and Control, no.8, pp. 338-353.*

Zadeh, L. (1978) - Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems, no.1, pp.3-28.*

Annexe 1

Synthèse sur la combinaison des représentations probabiliste et possibiliste

A.1 INTRODUCTION

Ce chapitre est destiné à donner des pistes pour l'utilisation conjointe d'informations statistiques (probabilistes) et d'informations subjectives imprécises (possibilistes) pour mener des calculs d'analyse de risque lorsque les valeurs de paramètres d'un modèle mathématique du phénomène étudié sont mal connues. L'aspect original de ce problème tient à cette présence conjointe de deux types d'incertitude, ce qui rend non seulement les calculs plus difficiles à mener, mais surtout, la méthodologie à adopter non triviale, voire non unique.

On s'intéresse plus particulièrement aux problèmes de la forme suivante : on calcule, à partir des données existantes mal connues, la valeur imprécise d'une certaine grandeur et on vérifie si cette valeur dépasse ou non un seuil fixé par des régulations (par exemple un seuil de pollution...). C'est le problème direct. On s'intéresse aussi au problème inverse, à savoir quel est le domaine de certains paramètres du modèle qui assure qu'on ne dépassera pas les seuils fixés, lorsque la connaissance des autres paramètres est imprécise.

Pour permettre d'avancer sur cette question, qu'on peut appeler « propagation de l'incertain hétérogène», encore très peu abordée dans la littérature, il convient de faire le point sur les liens entre les principales théories de l'incertain : probabilités possibilités, probabilités imprécises, fonction de croyance. La représentation possibiliste utilise les fonctions d'appartenance d'intervalles flous, qu'on peut voir comme un empilement d'intervalles de confiance emboîtés. Il existe des méthodes de passage entre les représentations, dont les justifications commencent à être mieux comprises, notamment entre probabilité et possibilités, deux théories de l'incertain qui sont au confluent des autres représentations, et qui sont aussi les plus simples donc les plus applicables dans un premier temps. Les sous-chapitre A.2 et A.3 introduisent les principales théories de l'incertain et les méthodes de passages entre représentations.

On peut distinguer deux classes d'approches pour la propagation de l'incertain hétérogène (problème direct):

1) On exprime toutes les données mal connues dans un format unique et on se ramène au cas homogène.

2) On préserve la spécificité des incertitudes et on propage en deux étapes homogènes. Se pose la question de l'ordre dans lequel on procède (probabiliste puis possibiliste, ou l'inverse). Il semble que certains choix de cet ordre mènent à des procédures plus exploitables que d'autres choix.

La première approche est celle adoptée traditionnellement par les statisticiens : tout paramètre mal connu est représenté par une distribution de probabilité, même si les sources d'informations sont de fait hétérogènes (mesures, plus ou moins entachées d'erreur, et opinions d'expert ; utilisation de probabilités uniformes ou Gaussiennes pour pallier l'absence de données). Les Bayésiens prétendent justifier cette approche sur

des bases « rationnelles » et en excluent toute autre. La technique de propagation la plus usuelle est la méthode par tirage aléatoire de Monte-Carlo. Une autre approche homogène existe pourtant : elle consiste à représenter toutes les données (numériques) par des intervalles, et à faire un calcul d'erreur. Cependant, le calcul d'erreurs ne repose pas sur les mêmes présupposés que la méthode de Monte Carlo. Dans cette dernière on fait souvent l'hypothèse d'indépendance entre les paramètres, ce qui crée des phénomènes de compensation des erreurs. En revanche, le calcul d'erreurs ne fait pas d'hypothèse d'indépendance (au sens stochastique), ni de dépendance, et s'avère être un calcul de pire et meilleur cas. C'est ce qu'on appelle souvent la non-interactivité entre les paramètres, l'idée étant qu'aucune configuration n'est exclue. La propagation possibiliste usuelle repose sur le calcul des intervalles flous, ce qui revient à effectuer le calcul d'erreurs sur les coupes de niveau des intervalles flous (qui sont des intervalles), et ce, pour tous les niveaux. Cette approche n'est pas beaucoup plus complexe que le calcul d'erreur usuel.

Quand on représente (via les transformations adéquates) toutes les données sous un format possibiliste, on a le choix entre les deux types d'hypothèses indépendance ou non-interactivité : cette dernière est admise dans le calcul d'intervalles flous usuel. Néanmoins il existe des calculs d'intervalles flous alternatifs (moins étudiés à ce jour, car plus complexes), dont un qui exploite l'idée d'indépendance (au sens de la compensation des erreurs). Si les données imprécises sont représentées par des fonctions de croyance, la méthode de propagation combinera calcul d'erreur et tirage aléatoire. Cette approche peut même simplifier la méthode de Monte Carlo usuelle dans certains cas, et fournir, à peu de frais, des encadrements des probabilités résultats. Ces questions sont abordées dans le sous-chapitre A.4. Le sous-chapitre A.5 discute de la comparaison d'un résultat incertain avec un seuil ou un objectif.

Pour ce qui est de la propagation hybride, on se restreint au cas où une partie des données est probabiliste, le reste étant possibiliste (représenté par des intervalles flous). Si on propage d'abord l'incertitude possibiliste, le résultat est un intervalle flou aléatoire (car dépendant des paramètres stochastiques). La question est alors d'extraire l'information pertinente et exploitable d'une statistique d'intervalles flous. Les variables aléatoires flous ont fait l'objet de nombreuses études (surtout théoriques) qui permettent de définir notamment une moyenne et une variance floues. Ces points sont abordés dans le sous-chapitre A.4.

Si on propage d'abord l'incertitude probabiliste, on obtient en théorie une distribution de probabilité sur la grandeur évaluée, indexée par les valeurs des paramètres possibilistes, ce qui semble très peu commode à représenter en pratique. On pourrait néanmoins faire un calcul d'intervalles aléatoires en effectuant pour chaque niveau de possibilité un calcul hybride d'intervalles et de Monte-Carlo. On obtiendrait alors une famille d'intervalles aléatoires indexé par des degrés de possibilité. Ce thème n'est pas abordé dans le rapport à ce stade, car il pose certainement des problèmes de complexité calculatoires et d'exploitabilité des résultats.

A.2. UN CADRE COMMUN POUR LES PROBABILITES ET LES POSSIBILITES

A.2.1. Mesures de Sugeno : définitions de base

On appelle capacité ou mesure de Sugeno (Sugeno, 1974) sur un référentiel Ω toute fonction telle que :

$\mu: 2^{\Omega} \rightarrow [0, 1]$

$A \subseteq B, \mu(A) \leq \mu(B)$ (monotonie dans l'inclusion)

Elle est dite normalisée si elle vérifie les propriétés :

$\mu(\varnothing) = 0 \qquad \mu(\Omega) = 1$

Ces propriétés sont minimales pour obtenir une fonction d'ensemble μ qui puisse permettre d'interpréter $\mu(A)$ comme l'évaluation d'un degré de confiance dans l'événement A.

Les mesures de probabilité, de possibilité, de nécessité, de plausibilité, de crédibilité sont des mesures de Sugeno. Elles sont respectivement définies par :

- Une capacité P est une mesure de probabilité ssi c'est une capacité additive c.-à-d. ssi : $\forall A, B \subseteq \Omega \; P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- Une capacité Π est une mesure de possibilité ssi :
  $\forall A, B \subseteq \Omega \; \Pi(A \cup B) = \max(\Pi(A), \Pi(B))$

- Une capacité N est une mesure de nécessité ssi :
  $\forall A, B \subseteq \Omega \; N(A \cap B) = \min(N(A), N(B))$

- Une capacité Pl est une mesure de plausibilité ssi elle est sous-additive à tout ordre n, c.-à-d. si,
  $\forall A_1, .., A_n \subseteq \Omega,$
  $Pl(A_1 \cup .. \cup A_n) \leq \sum_i Pl(A_i) - \sum_{i<j} Pl(A_i \cap A_j) + \ldots + (-1)^{n+1} Pl(A_1 \cap \cap A_n)$

- Une capacité Bel est une mesure de crédibilité (ou de croyance - « belief » en anglais) ssi elle est superadditive à tout ordre n, c.-à-d. si,
  $\forall A_1, .., A_n \subseteq \Omega,$
  $Bel(A_1 \cup .. \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \ldots + (-1)^{n+1} Bel(A_1 \cap \ldots \cap A_n)$

A l'ordre 2 par exemple, la sous-additivité et la super-additivité s'écrivent respectivement: $\mu(A \cup B) \leq \mu(A) + \mu(B) - \mu(A \cap B)$ et $\mu(A \cup B) \geq \mu(A) + \mu(B) - \mu(A \cap B)$. Il faut remarquer que la sous (resp. super)-additivité à l'ordre i

n'implique pas la sous-(resp. super) additivité à l'ordre i + 1, contrairement à l'additivité.

La mesure duale d'une capacité μ est la capacité μ' telle que :

$$\forall A \subseteq \Omega, \mu'(A) = \mu(\Omega) - \mu(\Omega / A)$$

Possibilités et nécessités, d'une part, et mesures de crédibilité et de plausibilité, d'autre part sont des fonctions duales l'une de l'autre : Si $\mu = \Pi$, alors $\mu' = N$, et si $\mu = Bel$, $\mu' = Pl$, et réciproquement.

Mesures de probabilité

Toute mesure de probabilité normalisée peut se définir à partir d'une distribution de probabilité $p : \Omega \rightarrow [0, 1]$ telle que $\sum_{\omega \in \Omega} p(\omega) = 1$ de la façon suivante :

$$P(A) = \sum_{\omega \in A} p(\omega)$$

p s'obtient évidement par : $\forall \omega \in \Omega, p(\omega) = P(\{\omega\})$

Les probabilités sont des mesures auto duales (la mesure duale d'une probabilité P est P elle même). On peut vérifier que toute probabilité est à la fois une plausibilité et une crédibilité.

Mesures de nécessité et de possibilité : théorie des possibilités

Toute mesure de possibilité (ou de nécessité) normalisée peut se définir, dans le cas fini, de façon équivalente à partir d'une distribution de possibilité $\pi : \Omega \rightarrow [0, 1]$ telle que $Max_{\omega \in \Omega} \pi(\omega) = 1$ de la façon suivante (Zadeh, 1978), (Dubois et Prade, 1987):

$$\Pi(A) = Max_{\omega \in A} \pi(\omega) \qquad N(A) = Min_{\omega \in \Omega / A} 1 - \pi(\omega)$$

π s'obtient évidement par : $\forall \omega \in \Omega, \pi(\omega) = \Pi(\{\omega\}) = 1 - N(\Omega \setminus \{\omega\})$.

Dans le cas infini (? = r), c'est faux. Il convient alors de partir de la distribution de possibilité π et de construire Π et N à l'aide des formules ci-dessus (remplaçant max et min par sup et inf, respectivement). π est alors la fonction d'appartenance d'un ensemble flou de réels, usuellement d'un intervalle flou caractérisé par le fait que les coupes de niveau $\{\omega: \pi(\omega) \geq \alpha\}$ sont des intervalles fermés pour tout $\alpha \in (0, 1]$
La mesure duale d'une possibilité est en effet une nécessité et réciproquement:

$$\Pi(A) = 1 - N(\Omega / A) \quad N(A) = 1 - \Pi(\Omega / A)$$

En autres propriétés des possibilités et des nécessités, on peut également vérifier que :

- $\Pi(A) < 1 \Rightarrow N(A) = 0$,

- $A \cup B = \Omega \Rightarrow \max(\Pi(A), \Pi(B)) = 1, \min(N(A), N(B)) = 0,$
- toute mesure de possibilité est une mesure de plausibilité ,
- toute mesure de nécessité est une mesure de crédibilité.

Enfin, on dit qu'une distribution de possibilité π_1 (ou les mesures de possibilité et de nécessité correspondantes) est plus spécifique que π_2 si π_1 est incluse dans π_2 au sens des ensembles flous, c'est-à-dire si :

$$\forall\ \omega \in \Omega,\ \pi_1(\omega) \leq \pi_2(\omega)$$

Mesures de plausibilité et de crédibilité (Shafer, 1976 ; Smets, Kennes, 1994)

Les mesures de plausibilité et de crédibilité normalisée peut se définir de façon équivalente à partir d'une distribution de poids m : $2^{\Omega} \rightarrow [0, 1]$ telle que $\sum_{A \in 2\Omega} m(A) = 1$ de la façon suivante :

$$Pl(A) = \sum_{B,\ A \cap B \neq \varnothing} m(B) \qquad Bel(A) = \sum_{B,\ B \subseteq A} m(B)$$

m est appelée une fonction de masse et tout A tel que $m(A) > 0$ est un élément focal : m(A) est la part de la croyance totale attribuée à l'affirmation « x ∈ A » exactement, et à aucune autre affirmation du type « x ∈ B ⊂ A ». Autrement dit, m(A) est la probabilité pour que la seule information disponible est « x ∈ A ». On note Foc(m) l'ensemble des élément focaux de m ($Foc(m) = \{F \subseteq \Omega, m(F) > 0\}$). La valeur m(?) est la probabilité qu'on ne sache rien (et vaut 1 quand on n'a pas d'information du tout). Dans ce cas Pl(A) = 1 et Bel(A) = 0 pour tout A ? Ø , ? . Entre autres propriétés des couples Bel, Pl on peut également vérifier que $Bel(A) \leq Pl(A)\ \forall\ A \subseteq \Omega$. Bel(A) évalue à quel point il est certain que l'information disponible, modélisée par la fonction m, implique « x ∈ A ». Pl(A) évalue à quel point l'information disponible ne contredit pas l'affirmation « x ∈ A ». Bel est plus exigeante que Pl : Bel (A) = 1 si « x ∈ A » est certain. Pl (A) = 1 si « x ∈ A » est seulement possible. En revanche, Pl (A) = 0 indique que « x ∈ A » est absolument impossible.

La mesure duale d'une plausibilité est une crédibilité et réciproquement. En effet :

$$Pl(A) = 1 - Bel(\Omega / A) \qquad Bel(A) = 1 - Pl(\Omega / A)$$

La fonction de masse m s'obtient depuis Bel en appliquant la transformation dite de Moebius. Il s'agit tout simplement de la solution du système d'équations linéaires :

$$\{Bel(A) = \sum_{B,\ B \subseteq A} m(B),\ \forall\ A \subseteq \Omega\},$$

dont les inconnues sont les m(B). On peut prouver qu'il y a toujours une solution positive ($m(B) = 0,\ \forall\ B$), pourvu que Bel satisfasse bien les axiomes des crédibilités.

Lorsque les éléments focaux de m sont des singletons, Bel(A) = Pl(A) pour tout A : les deux mesures sont alors des probabilités.

Lorsque, à l'inverse, les élément focaux de m sont emboîtés, Bel est une mesure de nécessité et Pl une mesure de possibilité (éléments emboîtés signifie que, pour toute paire d'éléments focaux A_i , $A_j \in Foc(m)$, $A_i \subseteq A_j$ ou $A_j \subseteq A_i$, ou, en d'autres termes, que les k éléments focaux de m forment une chaîne $A_1 \subseteq A_2 \subseteq ... \subseteq A_{k-1} \subseteq A_k$).

Notons que les fonctions de croyance ont été étudiées essentiellement dans le cas fini. Une fonction de masse sur les réels affectera donc des poids à un nombre fini de sous-ensembles de r, en pratique des intervalles. La notion de fonction de répartition (cumulée) devient une paire de fonctions de répartition supérieures et inférieures F* et F_* définies par $F^*(x) = Pl((-8, x])$ et $F_*(x) = Bel((-8, x])$ respectivement. On peut définir des fonctions de croyance continues dans des cas particulier, en utilisant des densités de probabilité bidimensionnelles continues $p(x, y) > 0$ si et seulement si $x = y$. Alors p(x, y) est interprété comme une « densité » de masse m([x, y]), et les fonctions de crédibilité et de plausibilité se calculent à l'aide des intégrales appropriées

A.2.2 Familles de probabilités inférieures et supérieures

Soit $\boldsymbol{P}$ une famille de mesures de probabilité sur un référentiel Ω. Pour tout $A \subseteq \Omega$, on peut définir :

sa probabilité supérieure	:	$P^*(A) = \sup_{P \in \boldsymbol{P}} P(A)$
sa probabilité inférieure	:	$P_*(A) = \inf_{P \in \boldsymbol{P}} P(A)$

En d'autres termes la valeur de la probabilité P(A) est imprécise:

$$\forall P \in \boldsymbol{P}, P_*(A) \leq P(A) \leq P^*(A)$$

L'ensemble $\boldsymbol{P^*} = \{P, P_*(A) \leq P(A) \leq P^*(A)\}$, appelé « enveloppe probabiliste », induit par les bornes issues de la famille $\boldsymbol{P}$ contient $\boldsymbol{P^*}$ mais $\boldsymbol{P}$? $\boldsymbol{P^*}$ en général, puisque le calcul de P^* et P_* est une projection de $\boldsymbol{P}$. Par exemple si famille $\boldsymbol{P}$ est un ensemble fini, $\boldsymbol{P^*}$ est un sous-ensemble convexe de $[0, '1]^{card(?)}$. C'est même un polyhèdre, car les contraintes qui définissent P sont linéaires $\boldsymbol{P^*} = \{P, P(A) \geq \alpha_A = P^*(A), \forall A\}$. La notion de fonction de répartition (cumulée) devient une paire de fonctions de répartition supérieures et inférieures F* et F_* définies par $F^*(x) = P^*((-8, x])$ et $F_*(x) = P_*((-8, x])$ respectivement.

On peut comprendre tout couple de mesures [Bel, Pl], et donc tout couple [N, Π], comme étant les probabilités inférieures et supérieures induites par une famille de probabilités :

- Bel (ou, de manière équivalente, Pl) définit la famille $\boldsymbol{P} = \{P, \forall A \subseteq \Omega, P(A) \geq Bel(A)\} = \{P, \forall A \subseteq \Omega, P(A) \leq Pl(A)\}$. Dans ce cas en effet $P_* = Bel$ et $P^* = Pl$ et

$\forall P \in \boldsymbol{P}$, $Bel \leq P \leq Pl$. Mais les probabilités inférieures (resp. supérieures) ne sont pas des crédibilités (resp. des plausibilités), en général[1].

- On peut comprendre tout couple de mesures [N, Π] comme étant les probabilités inférieures et supérieures déduites d'une famille définie à partir de $\boldsymbol{P} = \{P, \forall A \subseteq \Omega, P(A) \geq N(A)\} = \{P, \forall A \subseteq \Omega, P(A) \leq \Pi(A)\}$; : dans ce cas en effet, $P^* = \Pi$ et $P_* = N$. Dans le cas fini, $\boldsymbol{P}$ peut se définir à partir de π par : $\boldsymbol{P} = \{P, \forall i=1, k\ P(A_i) \geq 1 - \alpha_{i-1}\}$. où $A_i = \{\omega: \pi(\omega) \geq \alpha_i\}$.

- Réciproquement, étant donné $A_1 \subseteq A_2 \subseteq ... \subseteq A_{k-1} \subseteq A_k$ et $\alpha_i > 0$, i=1, k tel que $\boldsymbol{P} = \{P, P(A_i) \geq \alpha_i\}$, alors on sait que P^* est une possibilité et P_* une nécessité. Si les A_i ne sont pas emboîtés, P^* n'est pas forcément une plausibilité.

  Comme les intervalles [Bel, Pl] dont ils sont des cas particuliers, les intervalles [N, Π] sont donc les encadrement de probabilités mal connues : $N \leq P \leq \Pi$. Mais, puisque $\Pi(A) < 1 \Rightarrow N(A) = 0$, ce sont des intervalles généralement assez larges, de la forme [0, a] ou [b, 1]. On peut donc aussi définir une contrepartie de la fonction de répartition (cumulée) pour les mesures de possibilité, et de nécessité comme cas particulier des notions introduites plus haut : F^* et F_* sont définies (cas continu) par

$$F^*(x) = \Pi((-8, x]) = \pi(x) \text{ pour } x = \omega^*$$
$$= 1 \text{ sinon.}$$

et
$$F_*(x) = N((-8, x]) = 0 \text{ pour } x = \omega^*$$
$$= 1 - \pi(x) \text{ sinon}$$

où est ω^* une valeur telle que $\pi(\omega^*) = 1$.

A.3 TRANSFORMATIONS POSSIBILITES → PROBABILITES → FONCTIONS DE CROYANCE

Comme précisé plus haut, l'importance de ces transformations tient à la possibilité qu 'elles offrent de représenter toutes les données incertaines dans le même format.

A3.1. Probabilités → Fonctions de croyance

Aucune transformation n'est ici évidemment nécessaire, puisque toute probabilité est à la fois une plausibilité et une crédibilité : $Bel(A) = Pr(A) = Pl(A)\ \forall A \subseteq \Omega \Leftrightarrow$ les éléments focaux de m sont des singletons (soit : $m(F) > 0 \Rightarrow Card(F) = 1$)).

---

[1] On peut toujours appliquer la transformée de Moebius pour obtenir les solutions en m du système d'équations $\{\sum_{B,\ B \subseteq Ai} m(B) = P_*(A_i)\}$ – le système admet toujours une solution - et on peut vérifier que $P_*(A) = \sum_{B,\ B \subseteq A} m(B)$. Mais la fonction m ainsi calculée ne sera pas forcément positive sur tous les événements.

La fonction de masse correspondant à une probabilité coïncide avec la distribution de cette probabilité:

$$m(\{\omega_i\}) = p(\omega_i)$$

(y compris dans le cas continu).

A.3.2. Possibilités / Nécessités → Fonctions de croyance

Aucune transformation n'est nécessaire non plus ici, puisque qu'une mesure de possibilité (resp. de nécessité) est une mesure de plausibilité (resp. de crédibilité), plus particulièrement une mesure de plausibilité (resp. de crédibilité) dont les éléments focaux sont emboîtés.

Si l'on part d'une fonction de masse m à élément focaux emboîtés $A_1 \subseteq A_2 \subseteq ... \subseteq A_{k-1} \subseteq A_k$, $A_i \in Foc(m)$, la distribution de possibilité qui sous-tend la mesure est :

$$\forall\ \omega \in A_i / A_{i-1}, \pi(\omega) = \sum_{j=i,} m(A_j)$$

Ou, de manière équivalente : $\forall\ \omega \in A_i / A_{i-1}, \forall\ \omega' \in A_{i-1} / A_{i-2}$

$$\pi(\omega) = \pi(\omega') - m(A_{i-1})$$

Ainsi, les éléments de A_1 reçoivent une possibilité de 1, ceux de A_2 / A_1 une possibilité de $1 - m(A_1)$, ceux de A_i / A_{i-1} une possibilité de $1 - (m(A_1) + ... + m(A_{i-1}))$, etc. Dans le cas d'une fonction de croyance plus générale, on utilise la formule plus générale

$$\pi(\omega) = \sum_{\omega \in A} m(A).$$

Elle définit une fonction qui inclut distribution de possibilité et de probabilité.

Inversement, les éléments focaux de la distribution de masse codant une mesure de possibilité Π sont les couples de niveaux alpha de π ($1 = \alpha_1 > \alpha_2 > ... \alpha_k > \alpha_{k+1} = 0$):

$$A_i = \{\omega\ ,\ \pi(\omega) \geq \alpha_i\},\ \text{pour tout } \alpha_i \text{ t.q. } \exists\omega\ ,\ \pi(\omega) = \alpha_i$$

Et leurs masses sont données par :

$$m(A_i) = \alpha_i\ - \alpha_{i+1}\ (\text{donc } m(A_k) = \alpha_k \text{ pour l'élément focal le plus extérieur}).$$

Notons qu'on ne peut reconstruire complètement m à partir de π que dans le cas possibiliste (emboîté).

Dans le cas continu possibiliste, on peut définir la distribution de possibilité π à partir de deux fonctions f^- et $f^+ : [0, 1] \rightarrow r$, respectivement croissante et décroissante, telles que, $f^-(\alpha) = f^+(\alpha)\}$ et de la mesure Lebesgue λ (uniforme) sur [0, 1], à savoir $\pi(\omega) = \lambda(\{\alpha :\ f^-(\alpha) = \omega = f^+(\alpha)\})$. Dans ce cas, les coupes de niveaux $\{\omega: \pi(\omega) \geq \alpha\} = [f^-(\alpha), f^+(\alpha)]$.

A.3.3. Fonctions de croyances → probabilités

Le principe de la transformation dite « pignistique » par (Smets, 1990) , originellement introduite par (Dubois et Prade, 1982) et Williams (1982) est de généraliser le principe d'indifférence de Laplace aux fonctions de masse : prenons un m qui attribue toute la masse (= 1) à un seul élément focal F. Tous les éléments de F ont la même plausibilité.

D'après le principe de Laplace selon lequel ce qui est équiplausible est équiprobable, la distribution de probabilité correspondant à la fonction de masse m doit accorder à chaque élément de F la même probabilité 1 / Card(F). Appliquer ce principe à chacun des éléments focaux d'une fonction de masse quelconque revient à faire la somme convexe des probabilités obtenues pour chaque élément focal pris isolement, modulée bien sûr par les poids des éléments focaux m(F); ce qui justifie l'équation de la probabilité pignistique pp associée à une fonction de croyance.

$$pp(\omega) = \sum_{F,\, \omega \in F} (m(F) / Card(F))$$

La transformation pignistique a été formellement justifiée par Smets et Kennes (1990) dans le cadre des théories de représentation de l'incertitude, comme la façon rationnelle de parier à l'aide d'une probabilité subjective, pour un individu dont les connaissances sont fidèlement représentées par la fonction de masse m.

Les mesures de plausibilité et de crédibilité peuvent donc être comprises comme des enveloppes de probabilités chacune définissant un polyèdre $\boldsymbol{P} = \{P, P(A_i) \geq Bel(A_i)\}$ que l'on veut réduire à une probabilité unique. On doit choisir P dans ce polyèdre, afin de respecter la contrainte $P_*(A) \leq P(A) \leq P^*(A)$. La probabilité pignistique vérifie bien cette contrainte. Elle est même au centre de gravité de $\boldsymbol{P}$. C'est donc un choix respectant la symétrie du problème et reflétant au mieux les tendances exprimées par les paires (Bel(A), Pl(A)).

Elle a été en fait imaginée d'abord en théorie des jeux : si l'on considère que Ω est un ensemble de joueurs, et que Bel(A) rend compte de la force relative d'une coalition de joueurs, la transformation pignistique de Bel est l'indice de Shapley (XX REF) correspondant à cette capacité. $pp(\omega)$ évalue l'influence globale du joueur ω au travers de ses alliances. On peut replacer « joueur » par « critère », en théorie de la décision.

A.3.4. Fonctions de croyances → possibilités / nécessités

Dubois et Prade (1991) se sont posé la question de l'approximation d'une paire (Bel, Pl) par une paire (N, Π). On peut définir des approximations intérieures et des approximations extérieures d'un couple (Bel, Pl).

Approximation intérieure

Un couple (N, Π) définit une approximation intérieure de (Bel, Pl) ssi pour tout $A \subseteq \Omega$, $\Pi(A) \leq Pl(A)$ (ou, de manière équivalente, $Bel(A) \leq N(A)$). On cherche une approximation optimale qui maximise la largeur des intervalles [N(A), Π(A)]. On peut montrer qu'une seule distribution π satisfait cette exigence et elle est maximale au sens de l'inclusion d'ensemble flous. C'est la distribution la moins spécifique telle que $\forall A \subseteq \Omega$, $\Pi(A) \leq Pl(A)$ et elle coïncide avec le π défini plus haut par :

$$\forall\, \omega \in \Omega,\ \pi(\omega) = \sum_{A,\, \omega \in A} m(A) = Pl(\{\omega\})$$

Approximation extérieure

Un couple (N, Π) définit une approximation extérieure de de (Bel, Pl) ssi pour tout $A \subseteq \Omega$, $\Pi(A) \geq Pl(A)$ (ou, de manière équivalente, $Bel(A) \geq N(A)$). On cherche ici une approximation optimale qui minimise la largeur des intervalles [N(A), Π(A)]. Il n'y a plus unicité de la solution. On peut par exemple obtenir une approximation extérieure en ordonnant arbitrairement les éléments focaux A_i de 1 à n (n étant le nombre d'éléments focaux) et en posant :

$\Pi(A_1) = Pl(A_1)$,
$\Pi(A_1 \cup A_2) = Pl(A_1 \cup A_2)$, ...,
$\Pi(A_1 \cup \ldots \cup A_n) = Pl(A_1 \cup \ldots \cup A_n) = 1$.

Le problème est que, sauf cas particulier, il n'existe pas en général d'approximation extérieure la plus spécifique de l'intervalle [Bel, Pl]. Ces approximations on un sens pour des probabilités imprécises (P^*, P_*).

Un cas particulier où cet encadrement possibiliste extérieur optimal est unique est celui où les éléments focaux de m sont disjoints (par exemple, le cas où Bel = Pl est une probabilité) et linéairement ordonnés par m. Dans ce cas, la mesure de possibilité la moins spécifique telle que $A \subseteq \Omega$, $\Pi(A) \geq Pl(A)$ est la mesure Π définie par :

$$\Pi(A_{\sigma(1)} \cup \ldots \cup A_{\sigma(i)}) = Pl(A_{\sigma(1)} \cup \ldots \cup A_{\sigma(i)})\ \forall i = 1, n$$

où σ est la permutation des éléments focaux telle que $m(A_{\sigma(j)}) < m(A_{\sigma(i)})\ \forall\, i < j$

A.3.5. Possibilités / Nécessités → Probabilités

Puisque toute mesure de nécessité est une mesure de crédibilité particulière on pourra transformer une mesure de nécessité/possibilité en une mesure de probabilité en appliquant la transformation pignistique. Rappelons que le principe de la transformation pignistique est le principe de Laplace, qui dit que tout ce qui est équiplausible doit être équiprobable. La transformation pignistique d'une distribution de possibilité a été étudiée dans le cas discret (Yager 1982 ; Dubois, Prade 1982; Dubois et al 1993).

Appliquer les principes de la transformation pignistique revient donc à construire la distribution de probabilité à partir des coupes de niveaux alpha de la distribution et de la fonction de masse m déduite de π. En pratique si ? $= \{\omega_1, \omega_2, \ldots \omega_n\}$ et $\pi(\omega_1) = 1 = \pi(\omega_2) = \ldots = \pi(\omega_n) = \pi(\omega_{n+1}) = 0$)

$$pp(\omega_i) = (\pi(\omega_i) - \pi(\omega_{i+1}))/ i \text{ , pour } i = 1, n$$

En effet, si $\pi(\omega_i) > \pi(\omega_{i+1})$, alors $\{\omega_1, \omega_2, \ldots \omega_i\}$ est focal et $m(\{\omega_1, \omega_2, \ldots \omega_i\}) = \pi(\omega_i) - \pi(\omega_{i+1})$. Cette transformation est bijective.

La transformation pignistique dans le cas continu (Sandri 1991)

Dans le cas continu, cette transformation s'écrit :

$$p(\omega) = \int_0^{\pi(\omega)} (1/ \text{Card}((\pi)_\alpha) d\alpha \text{ où } (\pi)_\alpha = \{\omega, \pi(\omega) \geq \alpha\}$$

Par exemple, si π est un triangle isocèle de base a autour de 0, on obtient :

$$p(\omega) = \int_0^{\pi(\omega)} 1 / (2.\, a\, .\, (1 - \alpha)) \; d\alpha = -.\, (\log(\pi(\omega))/2a$$

Dans le cas où π est un intervalle flou de support [a, d], de noyau [b, c] et de type L/R ($\pi(\omega) = L(\omega)$ si $\omega \in [a, b]$, $\pi(\omega) = 1$ si $\omega \in [b, c]$, $\pi(\omega) = R(\omega)$ si $\omega \in [c, d]$), la transformation s'écrit :

$$p(\omega) = \int_0^{\pi(\omega)} 1 / (R^{-1}(\alpha) - L^{-1}(\alpha)) \; d\alpha$$

Par exemple, si π est trapézoïdal :

$L(\omega) = (\omega - a) / (b - a)$, $\quad R(\omega) = (\omega - d) / (c - d)$,
$L^{-1}(\alpha) = (b - a)\, \alpha + a$, $\quad R^{-1}(\alpha) = (c - d)\, \alpha + d$

Et l'on obtient :

$p(\omega) = \int_0^{\pi(\omega)} 1 / (p\, \alpha + q) \; d\alpha$ $\quad$ où $p = c - d - b + a$ et $q = a - d$
$p(\omega) = 1 / p\, .\, \ln((p\, \pi(\omega) + q) / q)$

Le maximum d'entropie (Klir, 1990)

Considérons une famille de probabilités ***P*** que l'on veut caractériser par l'un de ses éléments. L'objectif étant de biaiser au minimum l'information générée en réduisant ***P*** à une probabilité unique, Klir (1990) propose de caractériser cette famille par celui de ses éléments qui possède l'entropie maximale. Rappelons que l'entropie d'une mesure de probabilité est :

$$\text{Entropie}(P) = \Sigma_{\omega \in \Omega} p(\omega) \log(p(\omega))$$

Cette mesure est maximale pour les distributions uniformes et minimale en absence d'incertitude (i.e. quand $\exists\ \omega_0 \in \Omega$, $p(\omega_0) = 1$).

Le principe du maximum d'entropie va donc choisir dans ***P*** une probabilité maximisant $\Sigma_{\omega \in \Omega} p(\omega) \log(p(\omega))$. Rappelons que cette sélection revient à effectuer le maximum d'hypothèses d'indépendance possibles entre événements ; ce type de transformation est plus adapté aux probabilités objectives (statistiques) que subjectives. Notamment si ***P*** contient la distribution uniforme, c'est celle-là qui sera choisie aussi proche qu'elle soit du bord du polyèdre ***P***. Elle ne reflète dont pas du tout les tendances exprimées par les bornes de probabilités.

Quoiqu'il en soit, le principe de maximum d'entropie a été appliqué par Moral (1986) à la transformation d'une possibilité en probabilité. En considérant que les ω_i de Ω ont été ordonnées de manière à c e que $\pi(\omega_1) \leq \pi(\omega_2) \leq ... \leq \pi(\omega_n)$, la transformée de π est :

$$p(\omega_1) = \min_{i=1,n} \pi(\omega_i) / i$$
$$p(\omega_k) = \min_{i=k,n} (\pi(\omega_i) - \Sigma_{j=1,k-1} p(\omega_k)) / (i - k + 1)$$

A.3.6. Probabilités → Possibilités / Nécessités

Le principe ici est de trouver une approximation supérieure (extérieure) d'une probabilité P : on se retrouve donc dans un cas particulier de l'approximation supérieure de fonctions de croyances (cf. Section 3.3.2) : si on considère que P représente toute l'information disponible (issue d 'un histogramme par exemple), on cherche donc la possibilité Π la plus spécifique telle que $\forall A \subseteq \Omega$, $\Pi(A) \geq P(A) \geq N(A)$, afin de perdre le moins d'information possible. Le problème est qu'en général, n'y a pas unicité de la solution optimale, sauf si l'on ajoute que l'ordre donné sur ? par les distributions doit être conservé, c'est à dire que :

$$p(a) \geq p(b) \Leftrightarrow \pi(a) \geq \pi(b)$$

ce qui semble une condition raisonnable.

Approximation possibiliste extérieure d'une probabilité : cas discret

(Dubois, Prade 1982 ; Delgado, Moral 1987)

Dans le cas discret, la distribution π la plus spécifique satisfaisant aux conditions citées ci-dessus est :

$$\pi(\omega) = \Sigma_{\omega', p(\omega') \leq p(\omega)} p(\omega')$$

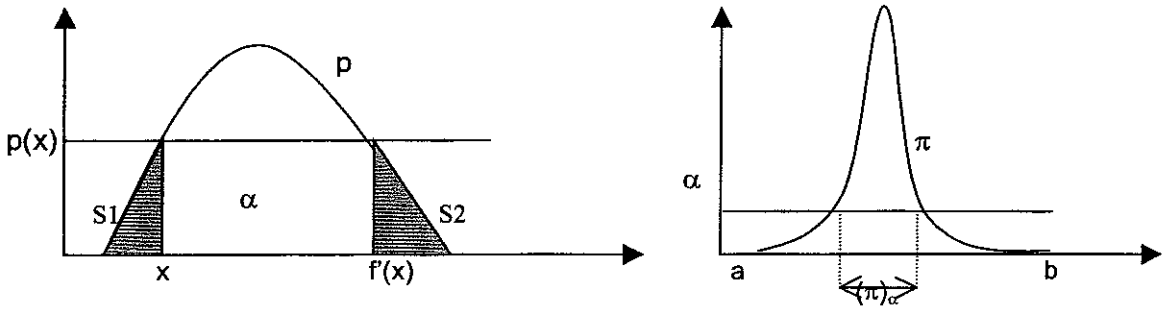
Si p est linéaire ($p(\omega_1) > p(\omega_2) > ... > p(\omega_n)$), on a : $\pi(\omega_1) = 1$, $\pi(\omega_2) = 1 - p(\omega_1)$, etc. On vérifie facilement que $\Pi(\{\omega_i, ..., \omega_n\}) = P(\{\omega_i, ..., \omega_n\})$. Cette transformation est bijective.

Approximation possibiliste extérieure d'une probabilité : cas continu (Dubois et al 1991)

Les mêmes principes s'appliquent si p est une distribution continue sur un sous ensemble de la droite réelle. Dans le cas d'une distribution continue unimodale (voir Figure ci-dessous) :

$$\pi(x) = S_1 + S_2 \text{ où } S_1 = P((-8\ x]),\quad S_2 = P([f'(x), +8))$$

avec $f'(x) = \max\{y, p(y) = p(x)\}$



Cela revient au calcul de l'intervalle de confiance [x, f'(x)] de probabilité α: tel que $\pi(x) = \pi(f'(x)) = 1 - \alpha = S_1 + S_2$ autour du mode de P. Dit autrement : $P((\pi)_\alpha) \geq 1 - \alpha$, où $(\pi)_\alpha$ est la coupe de niveau α de π.

Il faut noter que lorsque p est une distribution unimodale, π(x) est toujours une fonction concave de part et d'autre du mode. Si p est par exemple la distribution uniforme sur un intervalle [a, b], l'approximation extérieure symétrique la plus spécifique de p sera la distribution de possibilité triangulaire (a, (a + b) / 2, b)[2]. Cette distribution de possibilité est moins spécifique que la transformée de n'importe laquelle des probabilités de même support de densité unimodale symétrique. La distribution de possibilité triangulaire est donc un choix rationnel quand on ne connaît qu'un intervalle contenant le paramètre incertain, et qu'on suppose sa distribution symétrique.

Transformation pignistique inverse (Dubois, Prade 1983 ; Dubois et al 2001)

Les principes de la transformation pignistique d'une fonction de croyance en probabilité permettent d'imaginer une autre transformation possibiliste : on considère que p est en fait une probabilité subjective, qui est l'expression sous forme probabiliste d'une connaissance plus imprécise d'un agent (laquelle est supposée avoir la forme d'une fonction de croyance Bel). Ce ne sont que les règles imposées par le pari qui forcent l'agent à fournir une probabilité unique. On suppose qu'il choisit la transformation

[2] elle ne respecte pas la condition $p(a) \geq p(b) \Leftrightarrow \pi(a) \geq \pi(b)$. Si on l'imposait ici, on obtiendrait la possibilité uniforme (fonction caractéristique du support de p).

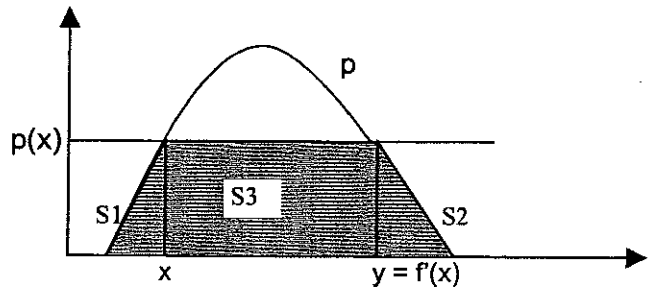
pignistique de Bel. En particulier si l'agent ne possède aucune information il fournit une probabilité uniforme.

Afin de modéliser correctement l'information détenue par l'agent, il faut donc rechercher la mesure de crédibilité dont p est la transformation pignistique. Comme il y a beaucoup de solutions, on fait l'hypothèse prudente d'ignorance maximale de l'agent. On cherche donc la moins informative des fonctions Bel de transformée pignistique p fournie par l'agent. On suppose que l'imprécision de Bel est mesurée par la cardinalité moyenne pondérée des éléments focaux ($I(Bel) = \Sigma\ m(A)\cdot Card(A)$). L'idée est que les grands ensembles sont plus imprécis que les petits.

On peut montrer que la mesure de crédibilité, de transformée pignistique p, qui maximise I(Bel) est toujours une mesure de nécessité ; plus précisément, quand p est une densité continue unimodale (voir figure ci-dessous), sa transformée pignistique inverse est définie par la distribution de possibilité π suivante :

$$\pi(x) = S_1 + S_2 + S_3$$

avec $S_1 = P((-8\ x])$, $S_2 = P([f'(x), +8))$ $S_2 = P([x, f'(x)])$,
où $f'(x) = \max(y, p(y) = p(x))$



Notons que si p est une distribution uniforme, la possibilité obtenue est elle aussi uniforme ($\pi(\omega) = 1$ pour tout $\omega \in [a, b]$).

Dans le cas discret, cette transformation s'écrit :

$$\pi(\omega) = \Sigma_{\omega' \in \Omega} \min(p(\omega), p(\omega))$$

A.4. SATISFACTION D'UN CRITERE PAR UNE FONCTION A PARAMETRES HETEROGENES : LE PROBLEME DIRECT

Soit une fonction f de k variables x_i dont certaines sont des variables aléatoires et d'autres des variables possibilistes et un critère sur f. On pourra par exemple prendre k = 2 et $f(x, y) = ax + y$, où x est contraint par une distribution de probabilité p et y par une distribution de possibilités π, et tester le critère $f(x, y) \leq e$. La question est de savoir dans quelle mesure le critère est satisfait.

A.4.1 Le cadre général de la présence simultanée de variables aléatoires et possibilistes

Lorsque toutes les variables de f sont des variables aléatoires, f est également une variable aléatoire dont on peut calculer la distribution de probabilité ou la fonction de répartition dans le cas discret:

$$P(f = u) = \Sigma_{u1, ..., uk,\ f(u1, ..., uk) = u}\ p(u_1, ..., u_k) = p_f(u).$$
$$P(f \leq e) = \Sigma_{u1, ..., uk,\ f(u1, ..., uk) \leq e}\ p(u_1, ..., u_k) = F_f(e)$$

où F_f est la fonction de répartition de $f(x_1, .., x_k)$. Cette dernière valeur est typiquement la probabilité que le critère soit satisfait .Si les variables aléatoires sont indépendantes :

$$P(f = u) = \Sigma_{u1, ..., uk,\ f(u1, ..., uk) = u}\ p_1(u_1) * ... * p_k(u_k).$$
$$P(f \leq e) = \Sigma_{u1, ..., uk,\ f(u1, ..., uk) \leq e}\ p_1(u_1) * ... * p_k(u_k).$$

On peut plus généralement calculer la probabilité de l'événement E :

$$P_f(E) = \Sigma_{u1, ..., uk,\ f(u1, ..., uk) \text{ satisfait } E}\ p(u_1, ..., u_k)$$
$$= \Sigma_{u1, ..., uk,\ f(u1, ..., uk) \text{ satisfait } E}\ p_1(u_1) * ... * p_k(u_k).$$

Dans le cas continu, on doit calculer la densité de probabilité ou la fonction de répartition avec les intégrales correspondantes aux formules ci-dessus, analytiquement si c'est possible, ou les approximer par différentes méthodes (par exemple une méthode de Monte Carlo).

Lorsque toutes les variables de f sont restreintes par des distributions de possibilité, les valeurs plus ou moins possibles de f seront également restreintes par une distribution de possibilité. Son calcul fera appel à des méthodes de calcul d'intervalles flous, analytiques si possible, ou à des approximations par alpha coupes.

$$\pi_f(u) = Sup_{u1, ..., uk,\ f(u1, ..., uk) = u}\ \pi(u_1, ..., u_k).$$

On peut également calculer l'équivalent des fonctions de répartition pour les mesures N et Π associées à la distribution, ou encore la possibilité et la nécessité que tel ou tel critère soit satisfait.

$$\Pi(f \leq e) = Sup_{u1, ..., uk,\ f(u1, ..., uk) \leq e}\ \pi(u_1, ..., u_k) = F_f^*(e).$$
$$N(f \leq e) = Inf_{u1, ..., uk,\ f(u1, ..., uk) > e}\ 1 - \pi(u_1, ..., u_k) = F_{f^*}(e),$$
$$\Pi(E) = Sup_{u1, ..., uk,\ f(u1, ..., uk) \text{ satisfait } E}\ \pi(u_1, ..., u_k),$$
$$N(E) = Inf_{u1, ..., uk \text{ ne satisfait pas } E}\ 1 - \pi(u_1, ..., u_k),$$

Si les variables possibilistes sont non-interactives (on ne fait alors pas d'hypothèse sur les liens entre les variables), on peut remplacer $\pi(u_1, ..., u_k)$ par $\min(\pi_1(u_1), ..., \pi_k(u_k))$ dans les formules précédentes, par exemple:

$$\pi(u) = \mathrm{Sup}_{u1, ..., uk,\ f(u1, ..., uk) = u} \min(\pi(u_1), ..., \pi(u_k))$$

Si les variables sont renseignées par des sources indépendantes , on peut remplacer $\pi(u_1, ..., u_k)$ par $\pi_1(u_1) \cdot ... \cdot \pi_k(u_k)$ dans les formules précédentes. Par convention, si π_i est la fonction d'appartenance d'un intervalle flou M_i, on note $f(M_1, ..., M_k)$ l' intervalle flou de fonction d'appartenance π_f.

La question est donc de savoir quelle est la nature du résultat lorsque f met en jeu les deux types de variables. Soient $x_1, ...x_i$ les variables aléatoires, et $x_{i+1}, ...x_k$ les variables possibilistes. Deux réponses sont possibles, selon la façon d'aborder ce problème.

Approche Hétérogène

Cette approche repose sur la remarque suivante : si les variables aléatoires $(x_1, ...x_i)$ ont leur valeurs fixées à $(u_1, ...u_i)$, $f(u_1, ...u_i , x_{i+1}, ...x_k)$ est une une variable possibiliste dont les valeurs sont restreintes par un ensemble flou $f(u_1, ...u_i , M_{i+1}, ...M_k)$. A chaque combinaison $(u_1, ...u_i)$ de valeurs des variables aléatoires correspond donc un ensemble flou, et une probabilité qui lui est attachée (la probabilité de la combinaison $(u_1, ...u_i)$) : $f(x_1, ...x_i , x_{i+1}, ...x_k)$ est un donc un nombre flou aléatoire décrit par une distribution de probabilité sur ... des ensembles flous !

En faisant une hypothèse de non interactivité sur les variables possibilistes, et supposant que x_j soit restreinte par l'intervalle flou M_i de distribution π_i pour $j = i+1,...,k$, la probabilité que la distribution de possibilité sur : $f(x_1, ...x_i , x_{i+1}, ...x_k)$ soit le sous ensemble flou M de Ω est :

$$p(f = M) = \Sigma_{u1, ..., ui,\ f(u1, ..., ui,\ Mi+1, ..., Mk) = M} p(u_1, ..., u_i)$$

On obtient donc une variable floue aléatoire.

Si tant est que le calcul de p est possible, on devrait alors calculer des distributions de probabilité sur les degrés de possibilité ou de nécessité de tel ou tel événement. Inversement, on devrait pouvoir calculer des intervalles de confiance à d% de ces mesures de possibilité.

En pratique, en appliquant une méthode de Monte-Carlo aux variables aléatoires on peut engendrer une « statistique floue » à partir d'un ensemble T de tirages de tuples $(u_1, ...u_i)$ c'est à dire un échantillon d'intervalles flous $\{f(u_1, ...u_i , M_{i+1}, ...M_k), (u_1, ...u_i) \in T\}$ qu'il convient d'exploiter.

Approche Homogène

Cette seconde approche considère que variables aléatoires comme variables possibilistes sont en fait des ensembles aléatoires décrits par des distributions de masse m_k, dont les éléments focaux U_k sont des singletons ou des intervalles consonants selon la nature de

la variable concernée. f est donc également un ensemble aléatoire, dont la fonction de masse est :

$$m_f(A) = \Sigma_{U1, ..., Uk, f(U1, ..., Uk) = A} m_1(U_1) * ... * m_k(U_k)$$

sous l'hypothèse d'indépendance des variables.

Le calcul de $f(U_1, ..., U_k)$ est effectué par le calcul d'erreurs. Il faut noter que les élément focaux $f(U_1, ..., U_k)$ de m_f ne sont en général ni emboîtés, ni des singletons. On obtient donc une fonction de croyance.

On peut à partir de cette distribution de masse calculer les indices :

$$Pl(f \leq e) = \Sigma_{A, inf(A) \leq e} m_f(A) = F_f^*(e),$$
$$Bel(f \leq e) = \Sigma_{A, sup(A) \leq e} m_f(A) = F_{f*}(e),$$
$$Pl(E) = \Sigma_{\forall a \in A, a \text{ satisfait } A} m_f(A),$$
$$Bel(E) = \Sigma_{\exists a \in A, a \text{ satisfait } A} m_f(A).$$

Là encore, le calcul analytique est souvent loin d'être simple dans le cas continu, en particulier quand la fonction met en jeu de nombreuses densités de probabilité. Néanmoins on peut approximer ces densités par des ensembles aléatoires FINIS dissonants, comme on le verra plus loin.

A.4.2 Calcul approché et exploitation des résultats pour l'approche hétérogène par Monte-Carlo plus analyse de sensibilité

Cette approximation repose sur une compréhension de f comme générant un ensemble flou aléatoire. Il s'agit d'effectuer une Monte Carlo sur les variables aléatoires seules pour estimer f à partir d'un échantillon d'ensemble flous. A chaque tirage, on fixe des valeurs pour chacune des variables aléatoires. Puisque ces variables sont devenues des constantes, on applique f à des constantes et des intervalles flous. On obtient à l'issue du Monte Carlo une famille d'intervalles flous (un par tirage) $\Phi = \{F_1, ..., F_n\}$: Φ est une approximation du nombre aléatoire flou $f(x_1, .., x_k)$. On cherche ensuite à estimer dans quel mesure cette approximation satisfait une critère e.

Dans Guyonnet et al (2001) il a été proposé de synthétiser cette famille en un ensemble flou unique F_d, tel que, dans d % des F_i, $\mu_{Fi}(x) \geq \mu_{Fd}(x)$. Pour ce, on envisage de procéder par alpha coupes : pour chaque niveau de coupe α, on possède donc une famille d'intervalles $(\Phi)_\alpha = \{(F_1)_\alpha, ..., (F_n)_\alpha\}$, où chaque $(F_i)_\alpha$, est de la forme $[b_inf_{i\alpha}, b_sup_{i\alpha}]$. On synthétise cette famille d'intervalles en un « intervalle d% » noté $[b_inf(\alpha, d), b_sup(\alpha, d)]$ et défini par :

$$b_inf(\alpha, d) = \sup \{ b_inf_{i\alpha}, Card(\{j : b_inf_{j\alpha} \geq b_inf_{i\alpha}\} \geq n * d / 100\}$$
$$b_sup(\alpha, d) = \inf \{ b_sup_{i\alpha}, Card(\{j : b_sup_{j\alpha} \leq b_sup_{i\alpha}\} \leq n * d / 100\}$$

b_inf(α, d) est la borne inf inférieure à d% des bornes infs et b_sup(α, d) est la borne sup supérieure à d% des bornes sup.

F_d est ensuite reconstitué à partir de ses coupes [b_inf(α, d), b_sup(α, d)]. F_d est donc une estimation « à d% », moyennant les approximations par Monte Carlo et par α coupes, des valeurs plus ou moins possibles de f. On peut donc calculer des estimations de la possibilité et de la nécessité que le critère soit satisfait. Par exemple :

$$\Pi_d(f \leq e) = \sup_{u \leq e} \mu_{Fd}(u) \qquad N_d(f \leq e)) = \inf_{u > e} 1 - \mu_{Fd}(u)$$

On peut aussi calculer des possibilité et des nécessité cumulées :

$\Pi_{d\,cumul}(e) = \Pi(f \leq e)$ est l'ensemble des point possiblement après f et
$N_{d\,cumul}(e) = N(f \leq e)$ est l'ensemble des point nécessairement après f

Cette méthode repose sur une hypothèse d'approximation à laquelle il faut faire attention : pour calculer les intervalles [b_inf(α, d), b_sup(α, d)], on traite des deux bornes indépendamment, alors qu'elles ne le sont pas : si d % des bornes inf sont supérieures à b_inf(α, d) et d % des bornes sup inférieures à b_sup(α, d), il est faux d'affirmer que d % des intervalles sont contenus dans les intervalles [b_inf(α, d), b_sup(α, d)]. Par exemple, considérons les intervalles suivants [1, 20], [2, 21], ..., [20, 39]. La borne inférieure à 95% est 2, la borne supérieure à 95% est 38. Or seuls 18 intervalles sur les 20 (i.e. 90%) sont compris dans [2,39]. Le problème est que les bornes inférieures et supérieures des intervalles sont traités comme des variables indépendantes alors qu'ils ne le sont pas. Cela implique que le F_d approximé par cette méthode ne peut pas être utilisé pour l'estimation de *tous* les types de critères. Heureusement, les critères ne mettant en jeu qu'une seule borne sont en général correctement estimés. En particulier, $N_d(f \leq e)$ estime correctement la nécessité que d% des intervalles flous satisfont le critère (ce calcul ne met en jeu que les bornes droites des intervalles) ; de même, $\Pi_d(f \leq e)$ estime correctement la possibilité que d% des intervalles flous satisfont le critère (ce calcul ne met en jeu que les bornes gauche des intervalles), et $[\Pi_d(f \leq e), N_d(f \leq e)]$ est bien un encadrement possibiliste de la probabilité que d% des intervalles flous tirés par Monte-Carlo satisfont le critère[3].

Autres méthodes possibles :

- Calculer la moyenne floue E(f) des $\{F_1, ..., F_n\}$:et un écart-type scalaire et(f) (voir Gil et al.). En déduire $N(E(F) \leq e)$ $\Pi(E(F) \leq e)$, $N(E(F) + et(f) \leq e)$ et $\Pi(E(F) + et(f) \leq e)$ $N(E(F) - et(f) \leq e)$ et $\Pi(E(F) + et(f) \leq e)$

[3] Le type de critère mettant en échec ce type de méthode est typiquement un critère mettant en jeu à la fois la borne gauche et la borne droite de l'intervalle flou, par exemple un critère de type $f \in [e1, e2]$.

• Calculer $N(f \leq e)$ et $\Pi(f \leq e)$ pour chaque $F \in :\{F_1, ..., F_n\}$. On obtient des statistiques $\{N_1(f \leq e), ..., N_n(f \leq e)\}$: et $\{\Pi_1(f \leq e), ..., \Pi_n(f \leq e)\}$, soit des histogrammes sur les degrés de satisfaction aléatoires du critère.

A.4.3. Approximation par une méthode de Monte-Carlo sur des ensembles aléatoires par l'approche homogène

Le principe de cette approximation est de considérer que les variables de f sont toutes de même type, à savoir des variables contraintes par des fonctions de croyances (puisque possibilités comme probabilités sont des fonctions de croyance) – f est donc un ensemble aléatoire. Puis on effectue une discrétisation sur ces fonctions de croyance, de manière à pourvoir appliquer un algorithme de Monte-Carlo.

Méthode de Monte-Carlo sur des fonctions de croyance

Lorsque les éléments focaux sont des intervalles ou des points et qu'il sont en nombre infini, les méthodes de Monte Carlo s'appliquent à partir de fonctions de croyance quasiment aussi simplement qu'à partir de distributions probabilité. En effet, il est facile d'effectuer des tirages aléatoires parmi les élément focaux d'une distribution de masse, en respectant la probabilité de tirage donnée par la fonction de masse (probabilité de tirer $A = m(A)$).

A chaque itération du Monte Carlo, on va donc tirer un intervalle I_{xi} par variable de x_i de f. On peut alors calculer l'image de ces intervalles par f. Dans le cas simple où f est continue et monotone dans chacun de ses argument, $f(I_{x1}, ..., I_{xk})$ est un intervalle, dont la borne inférieure est obtenue en appliquant f aux bornes inférieures des x_i croissants et aux bornes supérieures des x_i décroissants, et la borne supérieure de $f(I_{x1}, ..., I_{xk})$ est obtenue en appliquant f aux bornes supérieures des x_i croissants et aux bornes inférieures des x_i décroissants. A partir de N tirages, on obtient un ensemble de N intervalles : $\Phi = \{F_1, ..., F_N\}$ qui sont autant d'élément focaux d'une fonction de masse qui associe à chacun d'eux la masse 1 / N.

Cette approche permet de calculer approximativement une fonction de variables hétérogènes probabilistes et possibilistes continues. Pour les variables aléatoires on tire des valeurs au hasard selon la distribution; pour les variables possibilistes, on produit des coupes de niveau au hasard en tirant un niveau de possibilité au hasard selon une loi uniforme, et en calculant la coupe de niveau correspondante. On obtient $\Phi = \{f(u_1, \ldots u_i, (M_{i+1})_{\alpha i+1} \ldots (M_k)_{\alpha k}), (u_1, \ldots u_i, \alpha_{i+1}, \ldots \alpha_n) \in T\}$.

Notons la différence avec le cas précédent : ici, on tire une coupe de niveau différente pour chaque variable possibiliste. En revanche, la méthode hétérogène considère que c'est le même niveau de possibilité qui est choisi pour toutes les variables possibilistes, ce qui mène au calcul d'intervalles flous. Le calcul homogène proposé ici suppose une indépendance stochastique entre toutes les variables.

Par exemple, si l'on s'intéresse à la fonction $f(x, y) = a \cdot x + y$ et $E = \{f(x, y) \leq e\}$, avec des variables possibilistes traitées comme des densités d'ensembles aléatoires, on tire à chaque itération i deux intervalles (alpha coupes) : $[x_i, x_i']$ et $[y_i, y_i']$. Les F_i sont de la forme $[x_i + a \cdot y_i, x_i' + a \cdot y_i']$. On a donc deux courbes d'encadrement de la probabilité :

$Bel(f(x, y) \leq e) = Card(\{i, , x_i' + a \cdot y_i' \leq e\}) / N$
$Pl(f(x, y) \leq e) = Card(\{i, , x_i + a \cdot y_i \leq e\}) / N$
$PP((f(x, y) \leq e) = \sum_{i,\, xi + a \cdot yi \leq e \leq ,\, xi' + a \cdot yi'} (x_i' + a \cdot y_i' - e) / ((x_i' - x_i + a \cdot y_i' - a \cdot y_i)). N)$

Méthode directe dans le cas fini

Si les variables sont toutes représentées par des intervalles aléatoires $\{m_1, ..., m_k\}$avec un nombre fini d'éléments focaux U_i , alors il suffit de calculer $m_f(A) = \Sigma_{U1, ..., Uk\,,\, f(U1, ..., Uk) = A}\ m_1(U_1) * ... * m_k(U_k)$ sans passer par une méthode de MonteCarlo, puisque le nombre de calcul est fini. On obtient encore un ensemble de N intervalles : $\Phi = \{F_1, ..., F_N\}$, cette fois pondérés par $m_f(F_i)$.

Pour tout événement E, on peut alors calculer un encadrement de sa probabilité :

$Bel(E) = \Sigma_{\,Fi \subseteq E}\, m_f(F_i) \leq P(E) \leq Pl(E) = \Sigma_{\,Fi \cap E \neq \varnothing}\, m_f(F_i)$
En particulier $Bel(f \leq e) = \Sigma_{\,supFi = e}\, m_f(F_i)$ et $Pl(f \leq e) = \Sigma_{\,infFi = e}\, m_f(F_i)$.

Une autre solution est d'extraire une distribution pignistique de cette fonction de croyance :

$$pp(x) = \sum_{i,\, x \in Fi} m_f(F_i) \ / ((sup(F_i) - inf(F_i)))$$

et si FF est sa fonction de répartition, $FF(e) = PP((f(x, y) \leq e)$ est la probabilité de demeurer en dessous du seuil.

Discrétisation d'une fonction de masse

Lorsque les éléments focaux de m ne sont pas en nombre fini (par exemple si m est une probabilité continue sur un intervalle de la droite réelle), ou s'ils sont en trop grand nombre, on peut approximer m par un m' moins spécifique au sens de l'inclusion des ensembles aléatoires.

On dit qu'une fonction de masse m est incluse dans m' ($m \subseteq m'$), ou que m est plus spécifique que m' ssi les trois conditions suivantes sont vérifiées :

(i) $\forall A \in Foc(m)$, $\exists B \in Foc(m')$ t.q. $A \subseteq B$
(ii) $\forall B \in Foc(m')$, $\exists A \in Foc(m)$ t.q. $A \subseteq B$
(iii) il existe une matrice d'affectation W, d'entrées $A \subseteq Foc(m)$, $B \subseteq Foc(m')$ telle que :

$$A \not\subset B \Rightarrow W(A, B) = 0,$$

$$\forall A \in Foc(m), m(A) = \sum_{B, A \subseteq B} W(A, B)$$
$$\forall B \in Foc(m'), m'(B) = \sum_{B, A \subseteq B} W(A, B)$$

En d'autres termes, la masse m(A) d'un élément focal de m sera partagée entre un ou plusieurs sur-ensembles B de A qui seront les éléments focaux de m', et la masse m' d'un B ne sera que la somme des masses cédées par les A qu'il contient. Un cas particulier intéressant est celui où l'on regroupe simplement les éléments focaux de m en surensembles (c'est typiquement ce que l'on fait en discrétisant m) :

$$\forall A \in Foc(m), \exists ! B \in Foc(m') \text{ t.q. } A \subseteq B \text{ et } \forall B \in Foc(m'), m'(B) = \sum_{B, A \subseteq B} m(A)$$

Les distribution de masses liées par une relation d'inclusion ont la propriété remarquable suivante (Dubois, Prade 1991) :

$$m \subseteq m' \Rightarrow [Bel, Pl] \subseteq [Bel', Pl']$$

C'est à dire qu'en approximant m par une fonction de masse moins spécifique, on obtient une approximation extérieure de Bel et de Pl : $Bel' \leq Bel$, $Pl' \geq Pl$. De plus, si f est une fonction d'une variable x, et m_f l'image de m par f (un ensemble aléatoire tel que $m_f(A) = m(f^{-1}(A))$), alors

$$m \subseteq m' \Rightarrow m_f \subseteq m'_f$$

C'est à dire qu'en approximant m par une fonction de masse moins spécifique, l'image de cette dernière est encore une approximation supérieure de l'image de la première. Si calculer m'_f est plus facile que de calculer m_f, on peut facilement calculer une approximation « encadrante » de m_f.

Une application directe de cette propriété est que la méthode de Monte Carlo appliquée ci dessus à des fonctions de masse continues, donc à des variables aléatoires ou à des variables possibilistes décrites par des distributions continues peut être remplacée par une méthode directe, exacte, avec un nombre de calculs finis, pour obtenir une approximation extérieure du résultat. Il suffit de discrétiser les distributions de possibilité et de probabilité en en construisant des approximations finies extérieures.

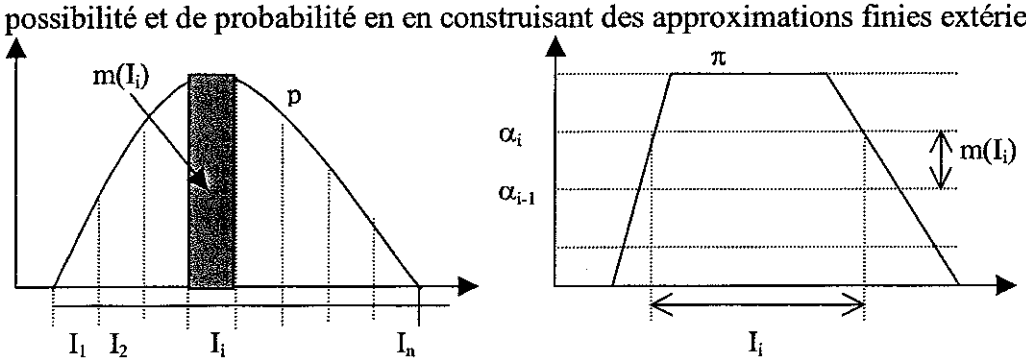


Fig. 6 - Discrétisation d'une distribution de probabilité, d'une distribution de possibilité

Discrétisation d'une distribution de probabilité: On va simplement partitionner le support [a, b] de la distribution de probabilité en n intervalles I_1, ..., I_n , typiquement des intervalles de même largeur $l = (b - a) / n$: $I_1 = [a, a + l]$, $I_2 = [a + l, a + 2 . l]$, ..., $I_n = [b - l, b]$ (on peut préférer partitionner [a, b] en intervalles de même probabilité). Ceci revient à transformer la fonction de masse m correspondant à p en une fonction de masse m' d'éléments focaux disjoints I_1, ..., I_n : $m(I_i) = P(I_i)$. Clairement, $p \subseteq m'$.
Discrétisation d'une distribution de possibilité: On va simplement extraire du support [a, b] de π n intervalles $I_1 \subseteq ... \subseteq I_n \subseteq [a, b]$, qui sont autant de coupes de niveau α_i de π. Typiquement, on peut prendre un $\delta = 1/n$ et fixer les α_i tous des δ : $\alpha_n = 1/n$, $\alpha_{n-1} = 2/n$, ..., $\alpha_i = \alpha_{i-1} + 1/n$, $\alpha_1 = 1$. On obtient donc la fonction de masse d'élément focaux $I_i = \{\omega, \pi(\omega) \geq \alpha_i\}$ et de masse $m'(I_i) = \alpha_i - \alpha_{i-1}$ (dans notre exemple de $m'(I_i) = \delta = 1/n$). Clairement, $\pi \subseteq m'$.

Principe d'un calcul homogène sur variables mixtes probabilistes et possibilistes

Résumons les principes exprimés ci dessus et appliquons les à notre fonction, $f(x, y) = ax + y$, où x est contraint par une distribution de probabilité p et y par une distribution de possibilités π :

- Discrétiser la distribution p en une fonction de masse m_x,
- Discrétiser la distribution en une fonction de masse m_y,
- Pour i = 1, N
  - Tirer selon m_x un élément focal $[x_i, x_i']$ de $Foc(m_x)$,
  - Tirer selon m_y un élément focal $[y_i, y_i']$ de $Foc(m_y)$,
  - Calculer $F_i = f([x_i, x_i'] , [y_i, y_i']) = [x_i + a . y_i, x_i' + a . y_i']$ et le mémoriser dans ***F***,
- Calculer : $Bel(E) = Card(\{F_i \in \boldsymbol{F}, F_i \subseteq E\}) \leq P(E) \leq Pl(E) = Card(\{F_i \in \boldsymbol{F} , F_i \cap E \neq \varnothing\})$

  C'est à dire : $Bel(f(x, y) \leq e) = Card(\{i, , x_i' + a . y_i' \leq e\}) / N$
  $Pl(f(x, y) \leq e) = Card(\{i, , x_i + a . y_i \leq e\}) / N$

Bien entendu, il n'est pas nécessaire de mémoriser tous les F_i et on peut imaginer un algorithme un petit peu sophistiqué qui calcule incrémentalement Bel(E) et Pl(E). On peut aussi simplifier la représentation finale en remplaçant deux éléments focaux A et B tres semblables par leur union affectée de la somme des masses.

Méthode homogène par transformation en possibilité ou probabilité

Une dernière façon d'aborder le problème serait de transformer les variables aléatoires en variables possibilistes, et d'effectuer le calcul d'une distribution de possibilité sur f, ou inversement, de transformer les distributions de possibilités en distributions de probabilité et d'effectuer un calcul de variables aléatoires.

Dans le premier cas, on irait vers une perte d'information, puisque l'on noie une probabilité dans une famille, mais l'on effectue un calcul sain puisque calculant des bornes peut être trop larges de la famille de probabilités sous-jacente. Mais on risque d'avoir des résultats plus imprécis que les autres méthodes, même si les calculs seront plus simples.

Le second calcul semble plus hasardeux, puis qu'il biaise la connaissance possibiliste pour la mettre sous forme de probabilité, et plus difficile (le calcul d'une fonction sur des variables aléatoires est en général plus coûteux qu'un calcul d'intervalles , fussent des intervalles flous) – mais il rend une information plus précise, une probabilité au lieu d'un intervalle de probabilité. C'est en fait l'option choisie par les statisticiens classiques qui supposent toutes les variables incertaines représentables par des distributions de probabilité. Cette option est méthodologiquement contestable (on néglige l'incomplétude des données) et pas très nouvelle.

Annexe 2

Article sous presse dans *Journal of Environmental Engineering*

A hybrid approach for addressing uncertainty in risk assessments

By: Dominique Guyonnet[1], Bernard Bourgine[1], Didier Dubois[2], Hélène Fargier[2], Bernard Côme[3] and Jean-Paul Chilès[1]

[1] : Envir. Spec., BRGM, BP 6009, 45060, Orléans Cédex 2, France
[2] : Math. Spec., Université Paul Sabatier, 31063 Toulouse, France
[3] : Envir. Spec., ANTEA, BP 6119, 45061 Orléans, Cédex 2, France

Key words: Risk assessment, Uncertainty, Monte Carlo, Fuzzy calculus

ABSTRACT: Parameter uncertainty is a major aspect of the model-based estimation of the risk of human exposure to pollutants. The Monte Carlo method, which applies probability theory to address model parameter uncertainty, relies on a statistical representation of available information. In recent years, other uncertainty theories have been proposed as alternative approaches to address model parameter uncertainty in situations where available information are insufficient to identify statistically representative probability distributions, due in particular to data scarcity. The simplest such theory is possibility theory, which uses so-called fuzzy numbers to represent model parameter uncertainty. In practice, it may occur that certain model parameters can be reasonably represented by probability distributions, because there is sufficient data available to substantiate such distributions by statistical analysis, while others are better represented by fuzzy numbers (due to data scarcity). The question then arises as to how these two modes of representation of model parameter uncertainty can be combined for the purpose of estimating the risk of exposure. This paper proposes an approach (termed a hybrid approach) which combines Monte Carlo random sampling of probability distribution functions with fuzzy calculus. The approach is applied to a real case of estimation of human exposure, via vegetable consumption, to cadmium present in the surficial soils of an industrial site located in the north of France. The application illustrates the potential of the proposed approach, which allows the uncertainty affecting model parameters to be represented in a way which is consistent with the information at hand. Also, because the hybrid approach takes advantage of the "rich" information provided by probability distributions, while retaining the conservative character of fuzzy calculus, it is believed to hold value in terms of a "reasonable" application of the precautionary principle.

INTRODUCTION

As risk assessments have become important aids in the decision-making process related to the management of sources of contamination, the issue of uncertainty with respect to model parameter values is of primary importance. Uncertainty affecting parameters in risk assessments may have essentially two origins: randomness (due to variability phenomena, or to the fact that all factors affecting the studied system cannot be modeled), and incompleteness (i.e., when information regarding parameter values is simply lacking). The ideal way to address uncertainty due to randomness is to collect data and perform a statistical analysis. When information is incomplete, or statistical data are not available, human experts can supply information on parameter values. Such subjective information can be tainted with various defects (ambiguity, vagueness, imprecision, etc.) and can be represented in various fashions (probability density functions, fuzzy numbers, uncertainty intervals, etc.). While different types of uncertainty may warrant different modes of uncertainty representation, the question arises as to how several modes of representation can be accommodated in the same estimation of risk ? Many researchers have addressed uncertainty using either one or the other of these modes of representation. For example Labieniec et al. (1997) used probability distribution functions to address uncertainty in the estimation of the risk of human exposure due to the presence of contaminated land. Prado et al. (1999) applied probability theory in risk assessments related to the underground disposal of nuclear waste. Dou et al. (1995), Bardossy et al. (1995), Freissinet et al. (1998), Cazemier (1999) present applications of fuzzy set theory (Zadeh, 1965) to environmental problems.

Comparatively few researchers have addressed the issue of combining different modes of representation of uncertainty in a single computational procedure (see for example Wonneberger et al., 1995). Yet this problem is of particular relevance in a risk assessment context, because in practice, information regarding the various parameters which influence the exposure to pollutants is not uniformly available. This paper proposes a method (here termed «hybrid») for combining probability distribution functions and fuzzy numbers in the same computation of risk. The method is first explained, and then applied to estimate the risk of human exposure to cadmium present in the surficial soils of an industrial site located in the north of France. It should be noted that the proposed method is one among several possible alternatives, the relative advantages of which are currently being investigated. The primary purpose of this paper is to promote a methodology that is consistent with the information at hand, and does not artificially add knowledge when it is actually lacking. As shown below, the proposed methodology might also be consistent with a "reasonable" application of the precautionary principle.

THEORY

Although it is not the purpose of this paper to provide a detailed presentation of the theories of probability and possibility, some basic principles are presented below for the sake of methodological clarity. For illustration purposes, we consider a variable X,

which will be represented on the one hand using probabilities, and on the other hand using possibility theory. If X is considered as a random variable, and the probabilities (P) that X should be less or equal than values x of X are known, a probability distribution function $F(x)$ can be defined as:

$$F(x) = P(X \leq x) \tag{1}$$

If variable X is continuous, i.e. it can take any value within a defined range, a probability density function $f(x)$ can be defined as:

$$f(x) = \frac{dF(x)}{dx} \tag{2}$$

These functions are illustrated graphically in Fig. 1(a), where a Gaussian probability distribution is assumed for variable X (in this case soil porosity). The distribution is defined by its mean (0.3) and its standard deviation (0.05). Ideally, in real situations, the definition of such a distribution would rely on the collection of a sufficient number of porosity measurements, so that meaningful statistical moments could be derived by fitting the theoretical distribution to a histogram of measured data relative frequencies. When dealing with risk assessments, however, it often occurs that for some parameters there are insufficient data to perform such a frequential analysis. In recent years there has been a large amount of research, especially from the nuclear industry, regarding how probability distribution functions can be derived using expert judgement (see for example Bonano et al., 1990). So-called "a priori", or subjective probability distribution functions can be defined using Bayesian probability theory (see for example Cullen and Frey, 1999). A fundamental postulate of this theory is that, whatever the available knowledge about some parameter, there exists a unique probability distribution which describes the available information for this parameter. This distribution is elicited from the expert(s) using different methods: either betting schemes, or the direct assessments of quantiles or confidence intervals with fixed levels of confidence (Cooke, 1991). In the estimation of risk, the probability distribution functions, whatever their origin, are sampled a large number of times in a random fashion using the Monte Carlo method (see for example Vose, 1996).

Some researchers (e.g. Wonneberger et al., 1995) have questioned the validity of deriving statistical moments or probability distribution functions from guesswork, even though based on expert knowledge. Expert knowledge is often incomplete, and the Bayesian setting often assumes more information than is actually available. For instance, under the Bayesian setting, it is impossible to distinguish the situation of complete ignorance (no information) from the situation of observed total randomness. The true strength of Bayesian theory lies in Bayes' theorem, which enables an *a priori* probability distribution function to be updated when data is subsequently collected. But in risk assessments, there is very seldom additional data to correct the initial assumption. Therefore, the use of prior probabilities may bring the user to artificially introduce knowledge that is not actually available, which can be misleading. Dubois and Prade (1992b) provide a critical discussion of the Bayesian approach to the modeling of expert opinions in risk assessments.

It would seem more cautious, and more faithful to the actual information supplied by experts, to assume partial knowledge of the actual probability distribution, if any. In principle, a family F of probability distributions can be substituted to the single Bayesian one. The intended meaning is that the actual probability function is one of them but the expert knowledge is insufficient to tell which one (Walley, 1991). In such a situation only partial probability evaluations can be obtained, in the form of probability intervals. The larger the intervals, the less informative. The difficulty with such an approach is that it is very unwieldy in practice. Possibility theory offers a very simple tool to encode special families of probability, which has been shown to be sufficient in practice (Dubois and Prade, 1992a; de Cooman and Aeyels, 1999).

According to possibility theory (Zadeh, 1978; Dubois & Prade, 1988), the available knowledge on a variable X can be represented by a "fuzzy number", which is characterized by its membership function (denoted μ) such that $\mu(x^*) = 1$ for some value x^*. This function, the values of which lie between zero and one, describes the "possibility" that variable X may take a certain value x. In this context it is called a possibility distribution. The term "possibility" refers to the idea of lack of surprise (Shackle, 1961); the more possible a value, the less surprising it is. The degree of possibility of an event A is denoted $\Pi(A)$ and defined as $\Pi(A) = Sup_{x \in A} \mu(x)$ where *Sup* designates the largest value. In probabilistic terms, a degree of possibility should be viewed as an upper bound of probability. Namely, considering the probability family F = {P, $P(A) = \Pi(A)$}, it holds that (Dubois and Prade, 1992a, De Cooman and Aeyels, 1999): $\Pi(A) = Sup_{P \in F} P(A)$. The dual function N of a possibility measure Π is called a necessity function, defined as $N(A) = 1 - \Pi(A^c)$ where A^c denotes the complement of A. Therefore a necessity degree can be viewed as a lower probability bound.

Figure 1(b) illustrates a fuzzy number for the same case as previously (soil porosity). Figure 1(b) informs us that it is considered most likely that porosity lies between 0.25 and 0.4, with no preference being expressed within this range, while values as low as 0.15 or as high as 0.5 cannot be totally excluded, and values outside this range are considered impossible. Note that although a probability density function could be defined with exactly the same shape as the membership function in Fig. 1(b), a fundamental difference between a fuzzy number and a probability density function is that in the former case, the integral below the function need not be equal to unity. In actual fact, a fuzzy number can be viewed as a pair of distribution functions respectively defined as:

$\Pi(X = x) = \mu(x)$ for $x = x^*$, and 1 otherwise (upper distribution),

$N(X = x) = 1 - \mu(x)$ for $x = x^*$, and 0 otherwise (lower distribution).

The area between these distributions reflects the incompleteness of the knowledge encoded in the fuzzy number: the wider the area the less information is present. An important particular case of a possibility distribution is the membership function of an interval [a, b], such that $\mu(x) = 1$ if $x \in$ [a, b], and 0 otherwise. The upper and lower distributions are then the distributions of Dirac functions in a and b respectively. It encodes all the probability distributions whose support is the interval [a, b].

Compared with a probability density function, a fuzzy number is "poor" in terms of intrinsic information. A probability density function defines a variable entirely, but applies best to systems which are "closed" (such as the throw of a dice). But the systems of interest in an environmental context are "open" (Oreskes et al., 1994), and possibility

theory may appear as an eligible alternative for representing some of the uncertainty pertaining to such systems. Possibility theory provides the mathematical basis for performing calculations with fuzzy numbers. It is a straightforward extension of interval propagation (see Dubois & Prade, 1988; Dubois et al., 2000). Computing the possibility distribution of a result obtained as a function $f(X, Y)$ of two parameters X and Y restricted by possibility distributions μ_X and μ_Y is achieved by a simple process. First, a threshold α of the membership function is selected, and the intervals $A_\alpha = \{x, \mu_X(x) = \alpha\}$ and $B_\alpha = \{x, \mu_Y(x) = \alpha\}$ are obtained. They are called α-cuts (cf. Fig. 2). Note that the membership function μ can be recovered from these α-cuts as $\mu_X(x) = Sup\{\alpha,\ x \in A_\alpha\}$. The second step computes $f(A_\alpha, B_\alpha)$ using standard methods in interval analysis (Moore, 1966). This is trivial if the function f is monotonic. This is done for each possibility value α (in practice for a suitable sampling of them). The possibility distribution $\mu_{f(X, Y)}(x)$ is recovered as mentioned above from the (nested) intervals $f(A_\alpha, B_\alpha)$.

With respect to the *a priori* choice of probability distribution functions rather than fuzzy numbers, even when faced with lack of information, a question which was addressed by Guyonnet et al. (1999) is: "what does it change in practice ?". An important consequence in an environmental or health context is that the *a priori* assumption of probability distribution functions (noted hereafter PDF's), without justification by available information, may lead to an unconservative estimation (minimization) of risk. In order to understand this, one must recall that the probability that two independent events, A and B, should occur simultaneously, is the product of the probabilities of both events. Therefore, during Monte Carlo random sampling, scenarios that combine low probability parameter values have all the less chance of being selected. If a very large number of iterations is used, these scenarios will be realized, but with very low relative frequencies. When the results of the Monte Carlo analysis are compared with an acceptance criterion (for example a reference dose), for a certain level of probability (for example 95%), these scenarios will be eliminated because they fall within the 5% range of outliers. Had model parameter uncertainty been considered in terms of possibilities rather than probabilities, these low-likelihood scenarios might not have been discarded, because fuzzy number calculus does not account for error compensation since it is based on interval analysis.

To better compare fuzzy interval analysis with Monte-Carlo simulation, note first that the α-cut A_α can be viewed as a confidence interval with confidence value $1 - \alpha$. Indeed the degree of certainty (necessity) that $x \in A_\alpha$ is generally $1 - \alpha$. Performing fuzzy interval analysis is like performing sensitivity analysis with confidence intervals, the same confidence value being assumed for all the fuzzy parameters. However, the confidence threshold is varied, not fixed. The reader will refer to Dubois and Prade (1991) for more details regarding the comparison between random variables and fuzzy variables.

In the event that a given model involves some parameters which are justifiably represented by PDFs (in particular because there is sufficient data to substantiate these PDFs), while others are considered to be more adequately represented by fuzzy numbers, a method should be derived to combine these two modes of representation of uncertainty in the estimation of risk. There are different ways by which this might be

achieved. For example Wonneberger et al. (1995) suggest using possibility/probability transformations (Dubois & Prade, 1993, 1992a). Another approach would be to treat both types of uncertainty within a single, more general framework, using the "belief functions" proposed by Shafer (1976) (see also Smets and Kennes, 1994). While different alternatives exist, the approach described below (called the "hybrid" approach) is believed to be more intuitively amenable to the working environmental engineer. The respective advantages (and degrees of validity) of different alternatives to this problem are currently being studied.

In order to document the hybrid approach, we will consider the estimation of a dose resulting from the exposure of a human target to soil pollutants. This dose is calculated using a "model", M, which is a function of a certain number of parameters:

$$Dose = M(P_1, \ldots, P_n, F_1, \ldots, F_m), \quad (3)$$

where M = model; $P_1, \ldots, P_n = n$ model parameters, each being represented by a PDF; $F_1, \ldots, F_m = m$ model parameters, each represented by a fuzzy number. Note that the model can also involve precise (i.e. "crisp") parameter values. Calculation of the dose is performed by combining the Monte Carlo random sampling technique, with the above-described method of α-cuts (Dubois and Prade, 1988) for fuzzy calculus. The hybrid procedure is summarized below, where steps have been indented in order to emphasize the iterative character of the calculations:

1. Generate n random numbers ($\chi_1, \ldots, \chi_n$) from a uniform distribution and sample the n PDF's to obtain a realization of the n random variables: $p_1, \ldots p_n$ (Fig. 3A)
   2. Select a value α of the membership function (a level of likelihood).
      3. Calculate the *Inf* (smallest) and *Sup* (largest) values of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$, considering all values located within the α-cuts for each fuzzy number (see Fig. 3.B).
      4. Assign these *Inf* and *Sup* values to the lower and upper limits of the α-cut of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$.
   5. Return to step 2 and repeat steps 3 and 4 for another α-cut (note: α can be increased stepwise from 0 to 1 every 0.1 increments). The fuzzy result of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$ (the fuzzy dose) is obtained from the *Inf* and *Sup* values of $M(p_1, \ldots, p_n, F_1, \ldots, F_m)$ for each α-cut.
6. Return to step 1 to generate a new realization of the random variables.

If steps 2 through 5 are repeated ω times, ω fuzzy doses are calculated (Fig. 3.C). For each value of the membership function (each value of α), the spread between the *Inf* and *Sup* values of the fuzzy results (see Fig. 3.C) is entirely a consequence of the Monte Carlo random sampling. It is therefore proposed to select the final *Inf* and *Sup* values of $M(P_1, \ldots, P_n, F_1, \ldots, F_m)$, for each value of α, by building a histogram of cumulative relative frequencies of the *Inf* and *Sup* values, and extracting the final *Inf* and *Sup* values for a certain level of probability. The final *Inf* and *Sup* values can be taken such that there is a 5% probability of respectively getting lower and higher values. This final fuzzy dose can be compared to an acceptance criterion, using tools provided by possibility theory, as illustrated in the next section. Note that in the general case, step 3

of the above procedure can be performed using a minimization and maximization algorithm. However, if the model is an equation involving simple monotonic functions, the *Inf* and *Sup* values can be identified directly.

One important consequence of combining probability distributions with fuzzy numbers is that the final result is a random fuzzy number (Gil, 2001): as the information conveyed by a fuzzy number is "poor" compared to a probability distribution, no mathematical procedure can compensate for this paucity and achieve the degree of system definition implied by a purely probabilistic representation. The proposed hybrid approach takes advantage of the "rich" information provided by the PDF's, but retains the conservative character of fuzzy calculus to account for those parameters for which a representation by unique PDF's is not justified by the available data. In the closing section, the consequences of assuming PDF's for all parameters are illustrated in the context of the following application example.

APPLICATION TO AN INDUSTRIAL SITE

The proposed method is applied to a metallurgical industrial site located in the north of France. The surficial soils of this site are contaminated by a number of metallic pollutants, including cadmium, which have been deposited by the smoke emanating from a chimney located on the site. The quality of the surficial soils (upper ten centimeters) has been monitored in detail, and there is a relatively large amount of data available. These data are primarily total soil metal contents measured by atomic absorption spectrophotometry after extraction by fluorhydric acid. No selective extraction data (see for example Tessier et al., 1979) were available for this study. This detail is of importance since we are interested here in the exposure of a human target to cadmium in the soil via the consumption of vegetables. A large body of scientific evidence shows that there is poor correlation between the amount of cadmium absorbed by plants, and the total amount present in the soil (see for example Jopony and Young, 1993, Lorenz et al., 1997). Metal uptake by plants depends on a variety of factors including metal speciation, plant specie, pH (Dijkshoorn et al., 1983, Singh et al., 1995), redox conditions, humidity, temperature (Chang et al., 1987), competition with other metals (Smilde et al., 1992, Chaney et al., 1999), etc. The variety of these factors explains why it is difficult in practice to develop a model of metal uptake by plants based on total soil concentrations, and therefore preferable to have direct site-specific measurement of metal uptake by plants. This is fortunately the case for this industrial site: measurements of cadmium content were performed (Luttringer & de Cormis, 1979) for a limited number of vegetables (in particular leeks), grown in the immediate vicinity of the soil sampling points. Measurements were performed on the edible vegetable parts, after they had been washed to eliminate the metal fraction present in the dust at the surface of the plant. Cadmium was analysed by atomic absorption after plant calcination and attack by chlorhydric and fluorhydric acid.

The first step in the analysis was to generate a map of soil cadmium content based on the measured data. Due to the relatively large number of surficial soil analyses (124), the spatial distribution of cadmium in the soils could be investigated using geostatistics (see Chilès and Delfiner, 1999). Geostatistics is a special branch of statistics, which applies to data which display a spatial structure. Its use is not compulsory for the hybrid

approach proposed in this paper, but served here to provide statistically representative estimators of soil cadmium concentrations outside measurement point locations. As could be expected, the data display a decrease in soil cadmium concentrations with increasing distance to the chimney (see Fig. 4 where Cd concentrations are in logarithm). In order to apply the classical geostatistical tools (the variogram and kriging), the data were first transformed into logarithm and then decomposed into a trend and a residual around this trend, according to:

$$\ln(\mathrm{Cd}_s) = \ln(\mathrm{Cd}_s)_T + R \qquad (4)$$

where $\ln(\mathrm{Cd}_s)$ = logarithm of measured soil cadmium concentrations; $\ln(\mathrm{Cd}_s)_T$ = predicted values of $\ln(\mathrm{Cd}_s)$ according to the trend; R = residual around this trend. A regression equation which describes the trend is:

$$\ln(\mathrm{Cd}_s)_T = -0.3 + 4.2\exp(-d/2) \qquad (5)$$

where d = distance to the chimney. The spatial distribution of the residual is then examined using the variogram. Directional variograms of the residuals reflected no anisotropy in the spatial distribution, which means that the effect of main wind direction is taken into account by the trend. Therefore an omnidirectional variogram was used. It was fitted with a linear variogram model and a nugget effect, which accounts for microstructures and/or measurement errors (Fig. 5). The next step consisted of using this variogram model to interpolate the residual R, by kriging, which ensures that the interpolation is not biased (on average the kriging error is zero) and that it is optimal (the kriging variance is minimum). Noting the kriged residual as R_K, we obtain the soil cadmium concentration from:

$$\mathrm{Cd}^*_{sM} = \exp(R_K + \ln(\mathrm{Cd}_s)_T) \qquad (6)$$

where Cd^*_{sM} is a median estimator of soil cadmium concentration since kriging of R yields a median estimator of R (i.e., the true value has a 50% probability of being lower than the kriged value). Results depicted in Fig. 6 show relatively high values of soil cadmium concentrations close to the chimney (up to around 25 ppm), and a gradual decrease to values below 1 ppm at a certain distance from the chimney. Since kriging achieves the minimization of the estimation variance, a by-product is the kriging variance or its square root, i.e., the kriging standard deviation σ_K. As we are concerned with the risk of exceeding a threshold value, we are interested in non-symmetrical confidence intervals (there is no risk associated with low values of soil cadmium content). If the kriging error (i.e., the difference between the kriged and true values) is assumed Gaussian, such confidence intervals can be deduced from σ_K. We write:

$$R < R_K + t\sigma_K \qquad (7)$$

where t is a factor which depends on the level of confidence assigned to R. For a level of confidence of 95%, for example, $t = 1.65$: R has 95% chances of being lower than R_K

+ 1.65 σ_K. An estimation of soil concentration is obtained by combining Equations (4), (6) and (7):

$$Cd_s < \exp(R_K + \ln(Cd_s)_T) \cdot (\exp(\sigma_K))^t \quad (8)$$

Since the first term in Equation (8) is the median estimator for Cd_s (Equation 6), $\exp(\sigma_K)$ can be considered as a multiplicative standard deviation. It is called here an "error factor" for short. The calculated error factor has values up to 1.3 with low points centered on measurement points. At these points, the remaining uncertainty is due to measurement errors. The statistical soil cadmium concentrations that result from this analysis are combined below with fuzzy uncertainties related to the uptake of cadmium by vegetables and to the absorption of a dose by a human target.

Measured values of cadmium in leeks are plotted in Fig. 7 as a function of measured soil concentrations at the locations where the leeks were grown. As seen in this figure, the small number of measurements (five) hardly supports a statistical analysis. They nevertheless provide very valuable information as they represent site-specific values of metal uptake by vegetables. This information can be analyzed using a fuzzy correlation. Based on the shape displayed by the measured data, it was chosen not to address plant uptake through the use of a "bioconcentration factor" such as appears in several risk analysis tools (for example HESP; Poels et al., 1990). Such a factor assumes that as soil concentration increases, plant uptake must increase proportionally. But agronomists argue that plant uptake cannot increase indefinitely, and that it will not exceed a certain level. Therefore a log-shaped correlation seemed more appropriate and consistent with the data depicted in Fig. 7. The following correlation equation is defined:

$$Cd_{pl}^* = Or + (As - Or)\left(1 - \exp(-k\, Cd_s)\right) \quad (9)$$

where Cd_{pl}^* = estimated cadmium concentration in the plant (mg Cd / kg dry plant); Cd_s = measured cadmium concentration in the soil (mg Cd / kg dry soil); Or = Cd_{pl}^* at the origin (Cd_s = 0); As = Asymptote (Cd_{pl}^* at large values of Cd_s); k = parameter which controls the rate of increase. Parameters Or, As and k in Equation (9) are represented by fuzzy numbers, while the Cd_s values are represented by probability density functions derived from the geostatistical analysis. In Fig. 7, the dashed lines (from Equation 9) represent "likely" boundaries for cadmium uptake by leeks. These boundaries imply that if someone were to go and measure the concentration in leeks cultivated on this site, it is likely that he/she would obtain a value that falls within these boundaries. The full line represents what is considered as a reasonably conservative upper boundary for cadmium concentration in leeks grown on this site. While one could argue regarding the precise positions of these curves, it should be noted that this representation is at least consistent with the measured data, and that it can be easily adjusted to accommodate input from agronomic experts. The fuzzy numbers for parameters Or, As and k are presented in Fig. 8 (along with another parameter discussed below). The limits of these fuzzy numbers are deduced directly from Fig. 7. For example As; the maximum cadmium content of the plant, is considered likely to be between 2.0 and 2.4 (these values have a likelihood of 1 in Fig. 8.B) while a value of 2.6 is considered as a possible upper boundary (in Fig. 8.B, the likelihood of values above 2.6 is considered nil).

The objective of the calculation is to estimate a dose of cadmium absorbed by a human target. In fact, doses are calculated along a grid which covers the site in order to examine the spatial distribution of absorbed dose. The dose is calculated from:

$$Dose = \frac{Cd_{pl}^{*} \cdot 1000 \cdot Con \cdot DMC}{BW} \qquad (10)$$

where Dose = Absorbed dose (μg Cd per day per kg body weight); *Con* = Leek daily consumption (kg leek per day); *DMC* = Leek dry matter content (weight percent); *BW* = Human target body weight (kg). The leek dry matter content (*DMC*) accounts for the fact that daily consumption is provided with respect to wet weight while the cadmium concentration in the leeks (Cd_{pl}^{*}) is relative to dry weight. The cadmium concentration in the leeks is obtained from the fuzzy correlation equation (Equation 9). As a simplifying hypothesis, it is considered that leeks are representative of vegetables with respect to human exposure to cadmium through vegetable consumption. The daily vegetable consumption and the vegetable dry matter content are selected based on data presented in INERIS (1999). Likely daily vegetable consumption is taken between 100 and 120 g/day, while an upper possible limit is taken as twice the higher value (240 g/day). The vegetable dry matter content (*DMC*) and the body weight are considered as constant: 15% and 70 kg respectively. It is assumed conservatively that the dose absorbed by the target is not influenced by vegetable preparation or assimilation.

An example calculation performed with the hybrid approach is presented for a median value Cd_{sM}^{*} = 7.97 ppm (from Fig. 6), and its corresponding error factor = 1.15. The probability distribution for Cd is obtained by calculating a normal Gaussian distribution for mean = ln(7.97) and standard deviation = ln(1.15), and then taking the exponential of the results to return to the distribution on Cd. The hybrid approach is performed according to Fig. 3 (which represents the general case). Note that as there is only one probabilistic variable involved, and because Equation (10) is very simple (minima and maxima can be identified directly), the calculation can also be performed without Monte Carlo random sampling and the use of minimization or maximization algorithms. Using the values for Cd_{sM}^{*} and the error factor above, the maximum value of Cd for a 95 % confidence level is obtained from Equation (8) for t = 1.65: Cd_{max} = 10.04 ppm. Likewise, the minimum value of Cd is obtained for t = -1.65: Cd_{min} = 6.33 ppm. These values then serve in Equations (9) and (10) to obtain the *Inf* and *Sup* values of the Dose, for each value of α (level of likelihood), using the relevant *Inf* and *Sup* values of the fuzzy numbers involved in Equation (10). The general hybrid approach, with Monte Carlo sampling, was also performed in order to check that results were identical. The results of this specific calculation are depicted in Fig. 9. Calculations were performed for a large number of points using the same grid as the one used to generate Fig. 6.

The next step consists in examining the acceptability of the calculated doses. This acceptability is considered here with respect to a maximum reference dose. According to WHO (1994), the kidney is the main target of cadmium toxicity. In order to maintain cadmium concentrations in the kidney cortex below 50 mg/kg, WHO (1994) recommends that cadmium absorption via food consumption should not exceed 1 μg per day and per kg body weight. The calculated fuzzy doses were compared to this daily

reference dose (noted D_o) using the degree of possibility (see Dubois and Prade, 1988) for the proposition: the calculated fuzzy dose F exceeds reference dose D_o. For a "crisp" reference dose, the degree of possibility is written:

$$\Pi(F > D_o) = \underset{u>D_o}{Sup}\, \mu_F(u) \qquad (11)$$

where $\mu_F(u)$ = membership function of F for any value u; Sup = the largest value. Fig. 10 provides a graphical illustration of Π for such a proposition. As long as the fuzzy dose is entirely below the reference dose, the possibility of D_o being exceeded is considered nil ($\Pi = 0$; Fig. 10a). As the fuzzy dose intersects the reference dose, excess of D_o is considered possible with a possibility degree $\Pi = \alpha$ (Fig. 10b). Once the reference dose intersects the plateau (Fig. 10c), Π becomes equal to 1. Note that the latter case does not imply that the reference dose will be exceeded with "certainty" (since Π refers to an upper probability bound only). Reasonable certainty occurs when another indicator of the validity of the proposition, i.e., the degree of necessity (Dubois and Prade, 1988), becomes equal to 1:

$$N(F > D_o) = 1 - \underset{u\leq D_o}{Sup}\, \mu_F(u) \qquad (12)$$

As stated previously, this degree can be viewed as a lower probability bound. Applying the possibility indicator to the calculated fuzzy doses, we obtain the spatial distribution of the possibility of exceeding the reference dose. This spatial distribution is depicted in Fig. 11. Possibilities of 0.45 of exceeding the reference dose are found in the close vicinity of the chimney, and decrease below 0.1 at a certain distance from the chimney.

DISCUSSION AND CONCLUSIONS

The map on Fig. 11 can serve as an aid in the decision-making process related to the management of this industrial site. It would be incumbent, however, upon the health authority, to define the acceptable level of "possibility" of reference dose violation. Depending upon the context, values around 0.1-0.2 may seem appropriate. From Fig. 11, such values would result in an area on the order of 20 square kilometers being ruled out for vegetable growth. To require a possibility of zero will in many cases be too strict, and result in excessive areas of land being ruled out for certain uses, or in excessive cleanup costs. It is recalled that the fuzzy calculus component of the proposed hybrid approach considers all possible combinations of fuzzy parameter values, and the calculated outcome does not result from any compensation of the uncertainties of these parameter values. As illustrated below, it is more conservative than a purely Monte Carlo calculation, and therefore constraints on acceptance criteria in terms of possibilities need not be as strict as in terms of precise probabilities resulting from a Bayesian analysis with assumptions of independence.

In order to illustrate the potential consequences of assuming probabilities throughout, the fuzzy numbers in Fig. 8 were converted to probability distribution functions, and hence to PDF's, by adjusting the y-axes so that the integrals beneath the curves equalled unity. Note that this procedure is not the mathematically rigorous manner for

establishing a correspondence between fuzzy numbers and probabilities (described in Dubois and Prade, 1993). It only mimics the *a priori* assumption of allocating probabilities rather than possibilities to the limit values. A high soil cadmium concentration was selected, i.e. a value from the immediate vicinity of the chimney (Cd_{sM}^{*} = 25 ppm, error factor = 1.2). The PDF's and the results of the Monte Carlo simulation (1 000 iterations) are shown in Fig. 12. In Fig. 13, the relative frequencies from the Monte Carlo simulation are compared with the reference dose. Based on this figure, the probability that the calculated absorbed dose should exceed 1 $\mu g/d\ kg^{-1}$ lies around 5%, and therefore the calculated dose would be considered as "tolerable". As the soil concentration value is located in the immediate vicinity of the chimney (highest values), thus the entire site would be considered suitable for vegetable growth. This calculation shows that if all parameters are assumed to be PDF's, the range of results is narrowed and insufficient weight is given to outlier scenarios which can have important implications in the decision-making process. On the other hand, if all parameters were represented by fuzzy numbers, despite the fact that some of them could be arguably represented by PDF's, the range of results would be too conservative. Thus the proposed hybrid approach is believed to hold some value in terms of consistency with a "reasonable" application of the precautionary principle.

Despite the potential advantages of the proposed hybrid approach, there are at least two important limitations regarding its application. Because in the general case (see Fig. 3), the approach combines two computationally-intensive methods (Monte Carlo sampling and fuzzy calculus by the method of α-cuts), its application would appear to be limited to calculations involving relatively simple models, such as analytical solutions, rather than numerical models requiring spatial discretization. There should be scope, however, for such an approach in a risk-assessment framework, as fate and exposure models often involve relatively simple equations.

The second, and probably more serious limitation, relates to the social acceptance of a level of possibility of risk violation (see the beginning of this section). Probabilities of risk violation have only recently gained some degree of acceptance by regulatory authorities. According to sociologists working in the field of industrial risks (Antoine-Paille, *pers. comm.*), in terms of social communication on risks, probabilities are very difficult to "sell" to the general public. In view of these difficulties, it is questionable whether possibilities of risk violation will be easily implemented in a regulatory framework. Casting possibility analysis in the framework of upper and lower probabilities may be the natural approach to addressing this issue. It is felt in any case that definition of possibility cut-off levels would require input from the field of social sciences.

In this paper a hybrid approach was proposed for combining probabilistic and possibilistic representations of model parameter uncertainty. As stated in Guyonnet et al. (1999), if data are available which substantiate a statistical representation of parameter value variability, then such a representation should certainly be preferred. Possibility theory is proposed as an alternative tool to assist in situations where such data are not available, and it is chosen not to force-fit probability distribution functions on data without statistical justification. Indeed, the subjective probability approach is not adapted when it is important to explicitly account for incomplete information. A subjective probability distribution may express a lack of information as well as the knowledge that the phenomenon is actually random. On the contrary, the hybrid

approach clearly distinguishes between the case where the precise value of a parameter is unknown because of its statistical variability, and the case where it is unknown due to a simple lack of observed data. The potential apparent drawback is that it may deliver poorly informative results in some situations, but in such a case it will highlight the need for collecting new information. The proposed hybrid approach presents the double advantage of preserving the strengths of statistical analysis, while providing the flexibility of the fuzzy approach when such an analysis is not substantiated by the data. It is proposed as a means to improve consistency between calculation hypotheses and available information.

As was stated previously, the hybrid approach is only one alternative among others for combining different modes of representation of uncertainty in a same calculation. Further research is needed in order to identify the pros and cons of each alternative. Another interesting issue that will be investigated in the near future, is the inverse problem. In the application example presented above, the hybrid approach was applied in a forward mode to evaluate the possibility of exceeding a tolerable reference dose. The inverse problem is: given a tolerable reference dose, and considering the uncertainty (probabilistic and possibilistic) affecting model parameter values, which residual soil concentration is such that the "possibility" of reference dose violation remains below the value fixed by the health authority ? This issue is of particular relevance to risk-based corrective action.

Acknowledgements

We thank Métalleurop France for kindly providing some of the soil data analyzed in this paper. We are also grateful to the anonymous reviewers, whose comments helped us to clarify our original manuscript. This study was funded in part by the French Ministry of the Environment, within the GESSOL Programme led by the French National Institute for Agronomic Research (INRA). Funding was also provided by the BRGM's Directorate of Research Activities.

APPENDIX I. References

Bardossy, A., Bronstert, A., and Merz, B. (1995). "1-, 2- and 3-dimensional modeling of groundwater movement in the unsaturated soil matrix using a fuzzy approach." *Advances in Water Resources*, 18(4), 237-251.

Bonano, E., Hora, S., Keeney, R., and von Winterfeldt, D. (1990). "Elicitation and use of expert judgement in performance assessments for high-level radioactive waste repositories." *Sandia National Laboratories* Report SAND89-1821, Albuquerque, New Mexico.

Cazemier, D. (1999). "Utilisation de l'information incertaine dérivée d'une base de données sols. (Use of uncertain information derived from a soil data-base)." *Ph.D. Thesis of the National School of Agronomy*, Montpellier (France). 170 pp.

Chaney, R., Ryan, J., Li, Y.-M., and Brown, S. (1999). "Soil cadmium as a threat to human health." *In*: M. McLaughlin and B. Singh (eds.) Cadmium in Soils and Plants. Kluwer Academic Press, Dordrecht, Netherlands.

Chang, A., Page, A., and Warneke, J. (1987). "Long-term sludge application on cadmium and zinc accumulation in Swiss chard and radish." *Journal of Environmental Quality,* 16, 217-221.

Chilès, J.-P., and Delfiner, P. (1999). "Geostatistics: Modeling Spatial Uncertainty." Wiley, New York.

Cooke R. (1991). "Experts in Uncertainty." Oxford University Press.

Cullen, A.C., and Frey, H.C. (1999). "Probabilistic techniques in exposure assessment: a handbook for dealing with variablility and uncertainty in models and inputs." New York Plenum Press, 333 pp.

de Cooman, G., and Aeyels, D. (1999). "Supremum-preserving upper probabilities." *Information Sciences*, 118, 173 - 212.

Dijkshoorn, W., Lampe, J., and Van Broekhovern L. (1983). "The effect of soil pH and chemical form of nitrogen fertilizer on heavy metal contents in ryegrass." *Fertilizer Research,* 4, 63-74.

Dou, C., Woldt, W., Bogardi, I., and Dahab, M. (1995). "Steady-state groundwater flow simulation with imprecise parameters." *Water Resour.. Res.,* 31(11), 2709-2719.

Dubois, D., and Prade, H. (1988). "Possibility theory." New York Plenum Press, 263 pp.

Dubois, D., and Prade, H. (1991). "Random sets and fuzzy interval analysis." *Fuzzy Sets and Systems,* 42, 87-101.

Dubois, D., and Prade, H. (1992a). "When upper probabilities are possibility measures." *Fuzzy Sets and Systems,* 49, 95-74.

Dubois, D., and Prade, H. (1992b). "On the relevance of non-standard theories of uncertainty in modeling and pooling expert opinions." *Reliability Engineering and Systems Safety*, 36, 95-107.

Dubois, D., and Prade, H. (1993). "On possibility/probability transformations." *In*: R. Lowen and M. Roubens (eds.) Fuzzy Logic: State of the Art. Kluwer Academic Press, Dordrecht, Netherlands.

Dubois, D., Kerre E., Mesiar R., and Prade, H. (2000). "Fuzzy interval analysis." *In:* D. Dubois and H. Prade (eds) Fundamentals of Fuzzy Sets, Kluwer , Boston, Mass., The Handbooks of Fuzzy Sets Series , 483-581.

Freissinet, C., Erlich, M., and Vauclin, M. (1998). "A fuzzy logic-based approach to assess imprecision of soil water contamination modelling." *Soil & Tillage Research,* 47, 1-17.

Gil M.A. (2001). "Fuzzy Random Variables." Special issue of *Information Sciences*, 133, nos. 1-2.

Guyonnet, D., Côme, B., Perrochet, P., and Parriaux, A. (1999). "Comparing two methods for addressing uncertainty in risk assessments." *J. Envir. Engrg.*, ASCE, 125(7), 660-666.

INERIS (1999). "Méthode de calcul des valeurs de constat d'impact dans les sols. (Method for calculating risk-based concentration limits in soils)." *INERIS Unpublished Report*, April 1999, Verneuil-en Halatte, France.

Jopony, M., and Young, S. (1993). "Assessment of lead availability in soils contaminated by mine spoil." *Plant and Soil*, 151, 273-278.

Labieniec, P. Dzombak, D., and Siegrist, R. (1997). "Evaluation of uncertainty in a site-specific risk assessment." *J. Envir. Engrg.*, ASCE, 123(3), 234-243.

Lorenz, S., Hamon, R., Holm, P., Domingues, H., Sequeira, E., Christensen, T., and McGrath, S. (1997). "Cadmium and zinc in plants and soil solutions from contaminated soils." *Plant and Soil*, 189, 21-31.

Luttringer, M., and de Cormis, L. (1979). "La pollution par les métaux lourds à Noyelles-Godault et ses environs (Pas de Calais). (Pollution by heavy metals at Noyelles-Godault and surrounding area, Pas de Calais)." *Unpublished report of the National Institute for Agronomic Research (INRA)*, Montfavet (France), 12 pp.

Moore R. (1966). "Interval Analysis". Prentice-Hall, Englewood Cliffs, NJ.

Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994). "Verification, validation, and confirmation of numerical models in earth sciences." *Science*, 263, 641-646.

Poels, C., Gruntz, U., Isnard, P., Riley, D., Spiteller, M., ten Berge, W., Veerkamp, W, and Bonyinck, W. (1990). "Hazard assessment of chemical contaminants in soil." ECETOC Report No. 40, ISSN-0773-8072-40. European Chemical Industry Ecology & Toxicology Center, Brussels, Belgium.

Prado, P., Draper, D., Saltelli, S., Pereira, A., Mendes, B., Eguilior, S., Cheal, R., Tarantola, S. (1999). "Gesamac: Conceptual and computational tools to tackle the long-term risk from nuclear waste disposal in the geosphere." *European Commission Report* EUR 19113 EN. Office for Official Publications of the European Communities, Luxemburg, 87 pp.

Shackle G. (1961) "Decision Order and Time in Human Affairs." Cambridge University Press, UK.

Shafer, G. (1976). "A Mathematical Theory of Evidence." Princeton University Press.

Singh, B., Narwal, R., Jeng, A., and Almas, A. (1995). "Crop uptake and extractability of cadmium in soils naturally high in metals at different pH levels." *Commun. Soil Sci. Plant Anal.*, 26(13&14), 2123-2142.

Smets, P. and Kennes, R. (1994). "The transferable belief model." *Artificial Intelligence,* 66, 191-234.

Smilde, K., Van Luit, B., and Van Driel, W. (1992). "The extraction by soil and absorption by plants of applied zinc and cadmium." *Plant and Soil,* 143, 233-238.

Tessier, A.P., Campbell, G.C., and Bisson, M. (1979). "Sequential extraction procedure for speciation of particulate trace metals." *Analytical Chemistry,* 51, 844-850.

Vose D. (1996). "Quantitative risk analysis - A guide to Monte-Carlo simulation modelling." Wiley, New York.

Walley P. (1991). "Statistical Reasoning with Imprecise Probabilities." Chapman and Hall.

WHO (1994). "Quality directives for drinking water. Volume 1: recommendations. 2nd Edition." *World Health Organisation.* Geneva, Switzerland. ISBN 92 4 254460 4, 202 pp.

Wonneberger, S., Kistinger, S., and Deckert, A. (1995). "Unbiased guess, a concept to cope with fuzzy and random parameters ?" *European Commission Report* EUR 16199 EN. Office for Official Publications of the European Communities, Luxemburg, 42 pp.

Zadeh L. (1965). "Fuzzy Sets." *Information and Control,* 8, 338-353.

Zadeh, L. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and Systems,* 1, 3-28.

APPENDIX II. NOTATION

The following symbols are used in this paper:

As = Asymptote for Equation (9);
BW = body weight;
Cd_{pl} = measured cadmium concentration in the plant;
Cd_{pl}^* = estimated cadmium concentration in the plant;
Cd_s = measured cadmium concentration in the soil;
Cd_{sM}^* = median estimator of cadmium concentration in the soil;
Con = leek daily consumption;
d = distance to the chimney;
D_o = daily reference dose;
DMC = Leek dry matter content (weight percent);
Inf = smallest value;
$F_1, \ldots, F_m$ = *m* model parameters each represented by a fuzzy number;
k = constant controlling the rate of increase in Equation (9);
$\ln(Cd_s)$ = logarithm of measured cadmium concentration in the soil;
$\ln(Cd_s)_T$ = value of $\ln(Cd_s)$ predicted by a regression equation (trend);
M = model;
Or = Cd_{pl}^* at the origin in Equation (9);
$P_1, \ldots, P_n$ = *n* model parameters each represented by a PDF;
PDF = probability density function;
R = residual;
Sup = largest value;
α = value of the membership function μ;
α-cut = all values of parameter *X* within shaded area in Fig. 2;
$\mu_F(u)$ = membership function of *F* for any value *u*;
χ = random number;
σ_K = kriging standard deviation for *R*;
ω = number of Monte Carlo iterations;
Π = possibility measure.

APPENDIX III. FIGURES

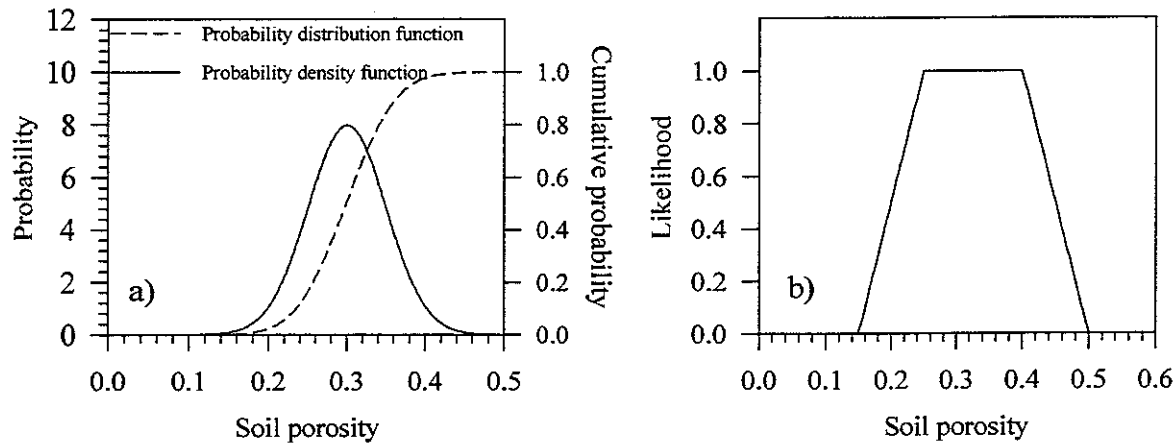


FIG. 1. Illustration of: (a) probability functions, and (b) a fuzzy number

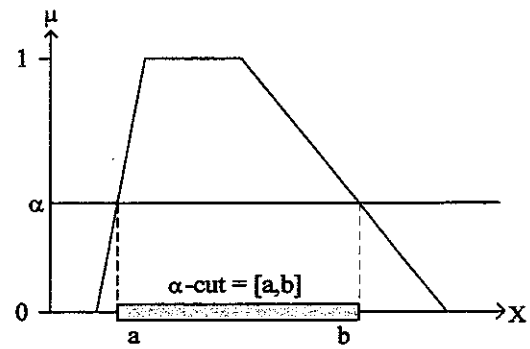


FIG. 2. Illustration of an α-cut; i.e., all values comprised between *a* and *b*

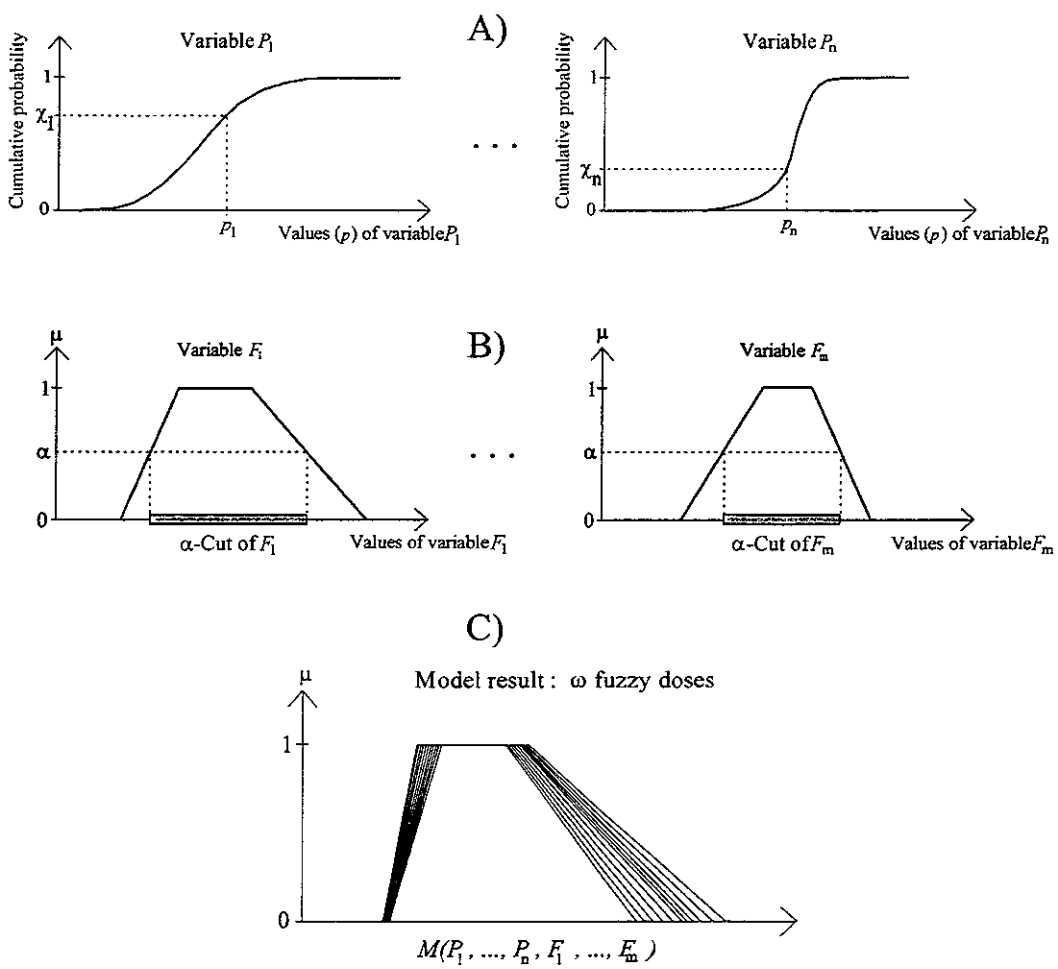


FIG. 3. Schematic illustration of the hybrid approach

A) Generate *n* random numbers (χ) to sample the *n* PDF's: obtain *n* values $p_1, ..., p_n$.
B) Apply method of α-cuts: look for *Inf* and *Sup* values of $M(p_1, ..., p_n, F_1, ..., F_m)$ on the α-cuts. Build fuzzy number of $M(P_1, ..., P_n, F_1, ..., F_m)$, reported in C).
C) Repeat the procedure ω times: obtain ω fuzzy results of $M(P_1, ..., P_n, F_1, ..., F_m)$

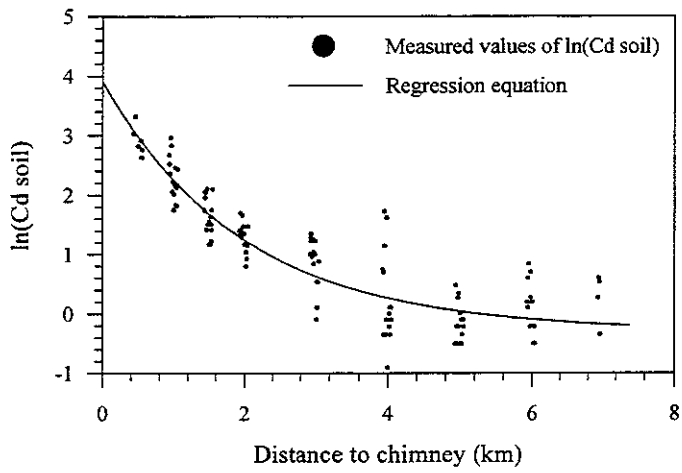


FIG. 4. Correlation between the logarithm of Cd soil concentrations, and distance between the chimney and the sampling point

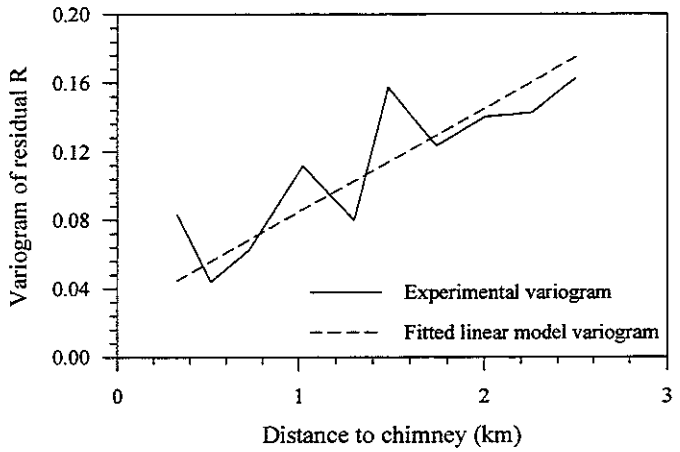


FIG. 5. Variogram of residual *R*. Fit with a linear variogram model

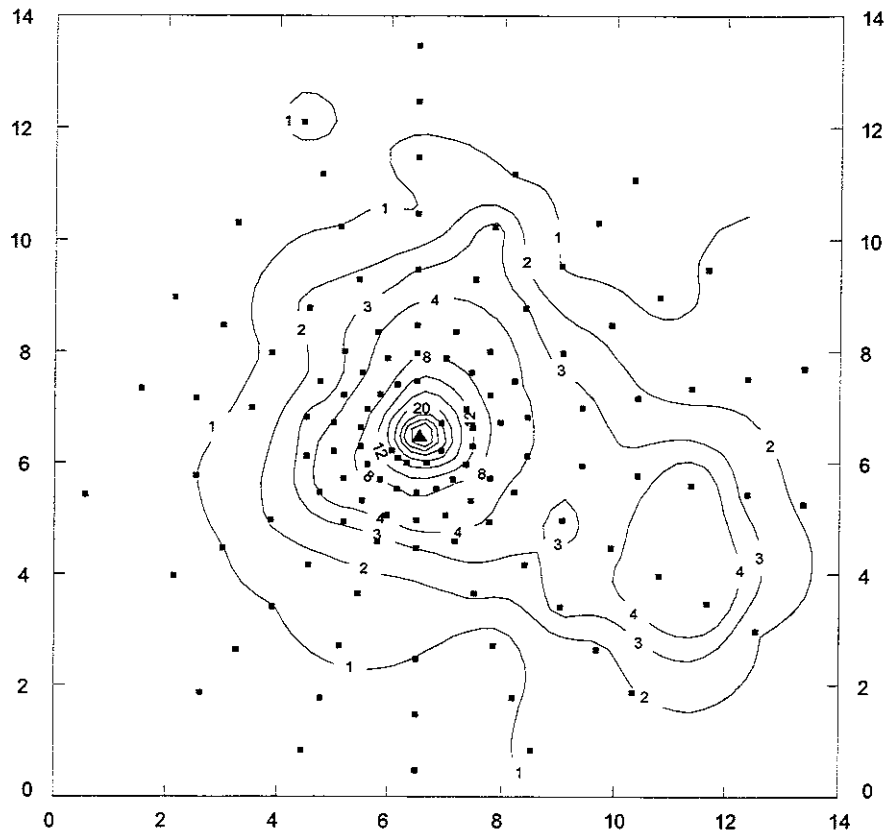


FIG. 6. Interpolated map of soil cadmium concentrations (Cd_{sM}^{*}; ppm). Graduation in km; Triangle = chimney location; Points = measurement points.

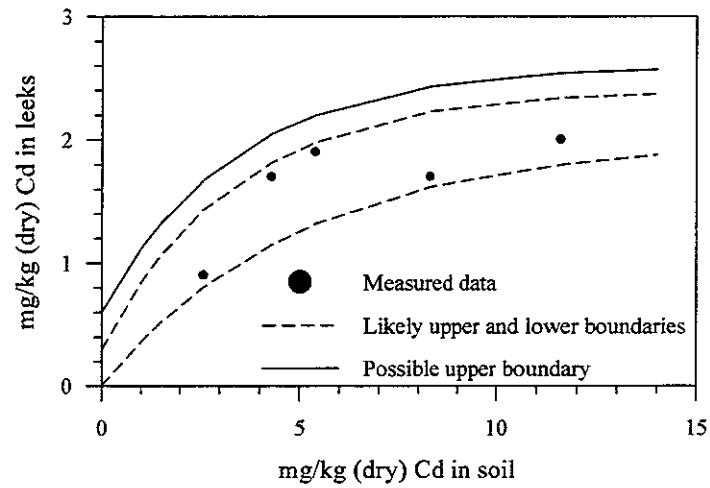


FIG. 7. Cadmium concentrations measured in leeks versus measured soil concentrations, and fuzzy correlation

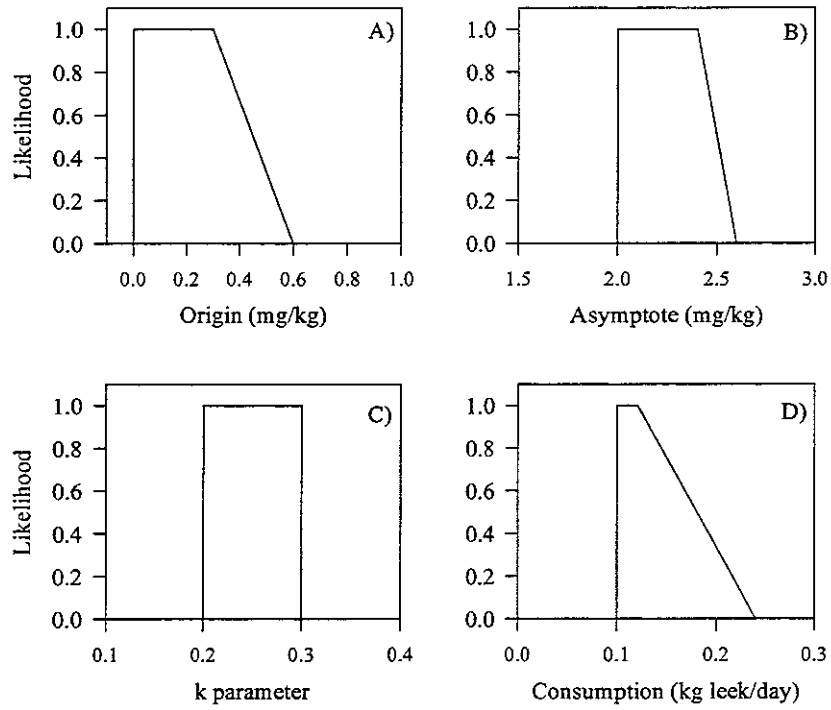


FIG. 8. Fuzzy numbers for several model parameters

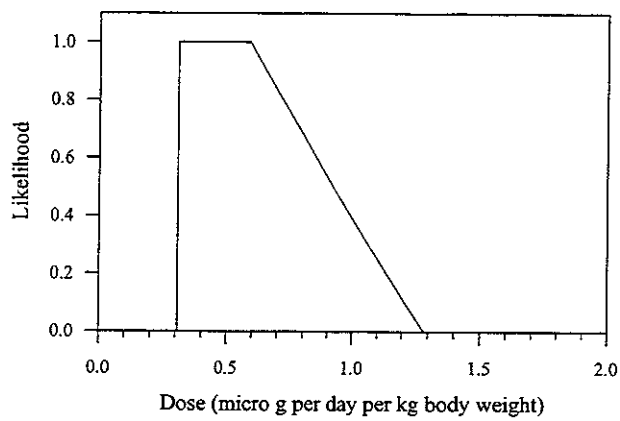


FIG. 9. Calculated dose for Cd_{sM}^{*} = 7.97 mg/kg, error factor = 1.15 and 95% confidence level on estimated soil cadmium concentration.

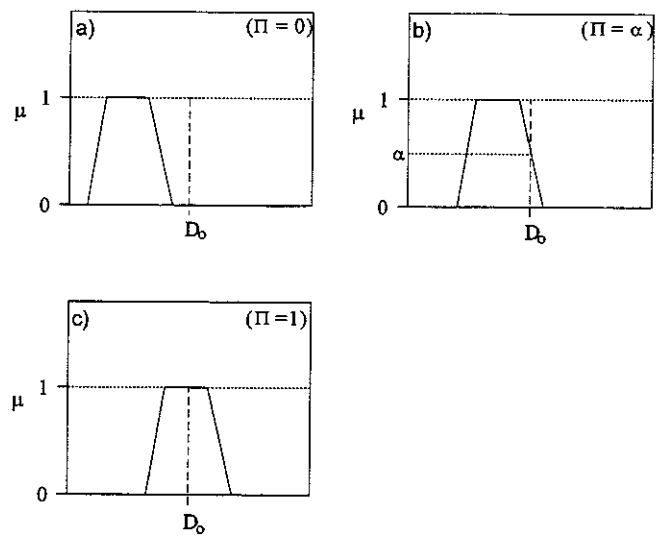


FIG. 10. Degree of "possibility" (Π) that a fuzzy dose should exceed a reference dose (D_o)

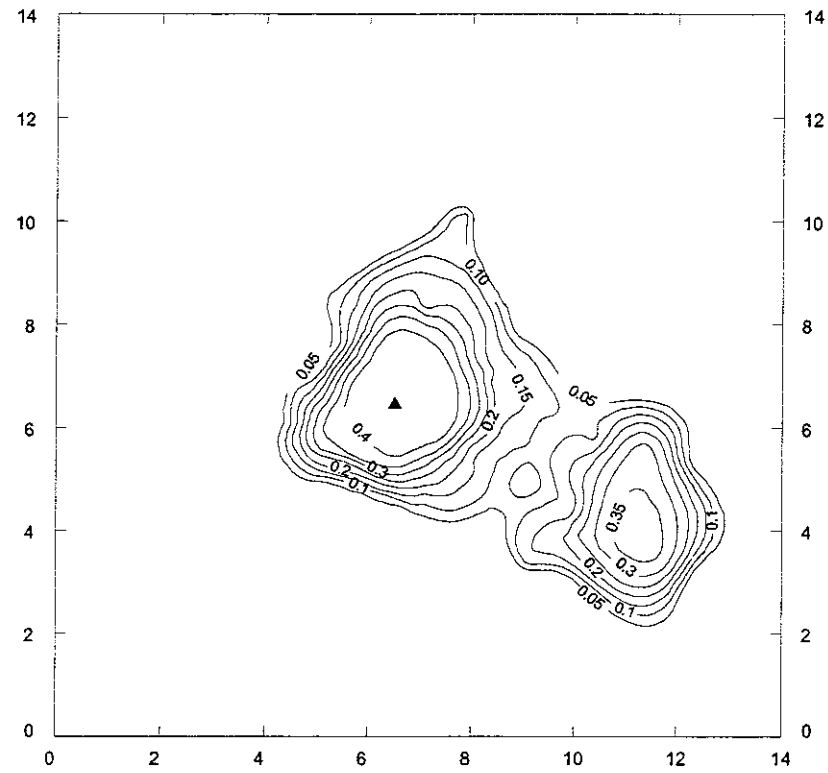


FIG. 11. Map of the "possibility" that the absorbed dose should exceed the reference dose (1 μg/d kg^{-1}). Graduation in km; Triangle = chimney location.

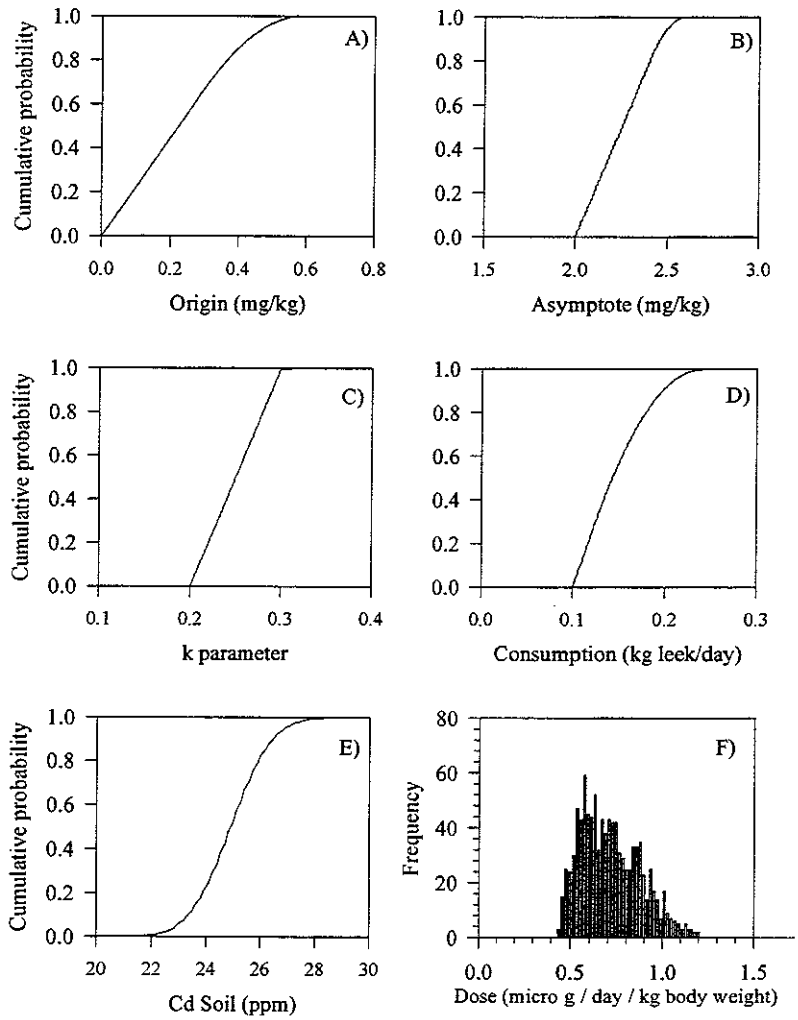


FIG. 12. A) through E): PDF's used in the Monte Carlo simulation. F): result of the Monte Carlo simulation (1000 iterations).

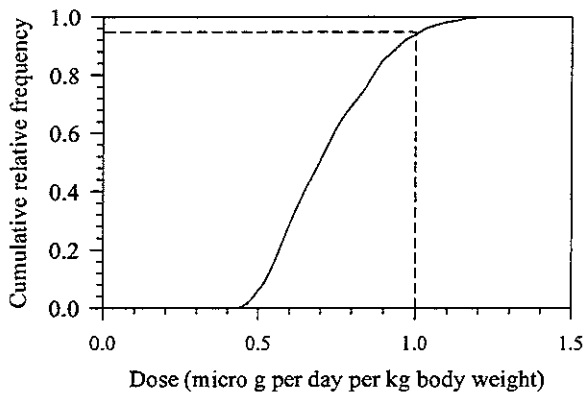


FIG. 13. Comparison between the dose distribution from the Monte Carlo simulation and the reference dose